FORM PTO-1390 (Modified) U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER	U.S. APPLICATION NO. (IF KNOWN, SEE 37 CFR
WIS4987P0051US	09/914001

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/US00/06456	10 MARCH 2000	11 MARCH 99 and 09 DEC 1999

TITLE OF INVENTION

Class II DNA Methyltransferases of Zea mays

APPLICANT(S) FOR DO/EO/US

Shawn M. Kaeppler, Nathan M. Springer, Michael G. Muszynski and Charles M. Papa

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☒ This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (24) indicated below.
4. ☐ The US has been elected by the expiration of 19 months from the priority date (Article 31).
5. ☒ A copy of the International Application as filed (35 U.S.C. 371 (c) (2))
   a. ☐ is attached hereto (required only if not communicated by the International Bureau).
   b. ☐ has been communicated by the International Bureau.
   c. ☒ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
   a. ☐ is attached hereto.
   b. ☐ has been previously submitted under 35 U.S.C. 154(d)(4).
7. ☒ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371 (c)(3))
   a. ☐ are attached hereto (required only if not communicated by the International Bureau).
   b. ☐ have been communicated by the International Bureau.
   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   d. ☒ have not been made and will not be made.
8. ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☒ An oath or declaration of the inventor(s) (35 U.S.C. 371 (c)(4)). (unsigned)
10. ☐ An English language translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371 (c)(5)).
11. ☒ A copy of the International Preliminary Examination Report (PCT/IPEA/409).
12. ☒ A copy of the International Search Report (PCT/ISA/210). (In the Information Disclosure Statement)

Items 13 to 20 below concern document(s) or information included:

13. ☒ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
14. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
15. ☐ A **FIRST** preliminary amendment.
16. ☐ A **SECOND** or **SUBSEQUENT** preliminary amendment.
17. ☐ A substitute specification.
18. ☐ A change of power of attorney and/or address letter.
19. ☒ A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825.
20. ☐ A second copy of the published international application under 35 U.S.C. 154(d)(4).
21. ☐ A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).
22. ☒ Certificate of Mailing by Express Mail
23. ☒ Other items or information:

Written Opinion
Response to Invitation to Furnish Nucleotide and Amino Acid Sequence Listing Complying with WIPOS Standard ST28
Express Mail Label No. EL904822474US

U.S. APPLICATION NO. (IF KNOWN, SEE 37 CFR)	INTERNATIONAL APPLICATION NO.	ATTORNEY'S DOCKET NUMBER
09/914001	PCT/US00/06456	WIS4987P0051US

24. The following fees are submitted:.

BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)) :		CALCULATIONS PTO USE ONLY
☐ Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO	$1000.00	
☐ International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO	$860.00	
☐ International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO	$710.00	
☐ International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4)	$690.00	
☐ International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4)	$100.00	
ENTER APPROPRIATE BASIC FEE AMOUNT =		$860.00
Surcharge of **$130.00** for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492 (e)).		$0.00

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	
Total claims	33 - 20 =	13	x $18.00	$234.00
Independent claims	19 - 3 =	16	x $80.00	$1,280.00
Multiple Dependent Claims **(check if applicable).**			☒	$270.00
TOTAL OF ABOVE CALCULATIONS =				$2,644.00
☐ Applicant claims small entity status. (See 37 CFR 1.27). The fees indicated above are reduced by 1/2.				$0.00
SUBTOTAL =				$2,644.00
Processing fee of **$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492 (f)). +				$0.00
TOTAL NATIONAL FEE =				$2,644.00
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31) **(check if applicable).** ☐				$0.00
TOTAL FEES ENCLOSED =				$2,644.00
			Amount to be: refunded	$
			charged	$

a. ☒ A check in the amount of **$2,644.00** to cover the above fees is enclosed.

b. ☐ Please charge my Deposit Account No. _____ in the amount of _____ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. **04-1644** A duplicate copy of this sheet is enclosed.

d. ☐ Fees are to be charged to a credit card. **WARNING:** Information on this form may become public. **Credit card information should not be included on this form.** Provide credit card information and authorization on PTO-2038.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

MUELLER, Lisa V.
Rockey, Milnamow & Katz, Ltd.
Two Prudential Plaza
180 N. Stetson Ave., Suite 4700
Chicago, Illinois 60601
312-616-5400 (Telephone)
312-616-5460 (Fax)

SIGNATURE

MUELLER, Lisa V.
NAME

38,978
REGISTRATION NUMBER

August 20, 2001
DATE

CLASS II DNA METHYLTRANSFERASES OF ZEA MAYS

FIELD OF THE INVENTION

The present invention relates to nucleic acid and amino acid sequences which encode class II DNA methyltransferases. The present invention further relates to methods of using the nucleic acid and amino acid sequences described herein to stabilize transgene expression in transgenic plants, to alter the yield or biochemical qualities of plants and to silence targeted genes in plants *in vivo*.

BACKGROUND OF THE INVENTION

The information content of a primary DNA sequence can be enhanced by the addition of a methyl group to the ring structure of cytosine or adenine residues (Finnegan, E.J., et al., *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49:223-47 (1998)). The chemical modification of DNA is known to affect protein-DNA interactions. Specifically, in prokaryotes, methylation of DNA prevents cleavage by the cognate restriction endonucleases. *Id.* In higher eukaryotes, cytosine methylation can inhibit binding of regulatory proteins and methylation of promoter and coding sequences of genes can repress transcription, both *in vitro* and *in vivo*. *Id.* Methylation of DNA has been implicated in the timing of DNA replication, in determination of chromatin structure, in increasing mutation frequency, as a causal agent for some human diseases, and as a basis for epigenetic phenomena. *Id.*

Eukaryotic genomes are not methylated uniformly, but instead contain specific methylated regions, with other domains remaining unmethylated (Martienssen, R.A., et al., *Current Opinion in Genetics and Development*, 5:234-242 (1995)). The enzymes that transfer methyl groups to the cytosine ring are cytosine-5-methyltransferases (hereinafter referred to as "DNA methyltransferases") and have been characterized from a number of eukaroytes. All characterized eukaryotic DNA methyltransferases exhibit little primary sequence specificity *in vitro* other than the short canonical symmetrical sites methylated which are CpG in animals, and CpG and CpNpG in plants (where N stands for any nucleotide). Mammalian and plant

genomes contain methylation-free GC-rich zones, or CpG islands, which are frequently associated with the 5' regions of housekeeping genes. *Id.*

In plants, DNA methylation is necessary for normal development. For example, Arabidopsis having reduced levels of DNA methylation demonstrate a range of abnormalities, including loss of apical dominance, reduced stature, altered leaf size and shape, reduced root length, homeotic transformation of floral organs and reduced fertility (Finnegan, E.J., et al., *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49:223-47 (1998)). Moreover, Arabidopsis plants in which methylation had been reduced by at least 70% became infertile after four to five generations of selfing. *Id.* A comparable reduction in DNA methylation is embryo lethal in mammals. *Id.*

Two classes of DNA methyltransferase enzymes have been cloned in plants (Finnegan, E.J., et al., *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49:223-47 (1998)) - class I and class II. Class I enzymes include MetI and MetII from Arabidopsis (Finnegan et al. *Nucleic Acids Res.*, 21(10):2383-2388 (1993); Nebendahl, et al., *Gene* 157(1-2):269-272 (1995)), Met1-5 and Met2-21 from carrot (Bernacchia, G et al., *Plant Physiol.* 116:446-446 (1998)), C-5 MTase from tomato (Bernacchia, G et al. *Plant J.*, 13(3):317-330 (1998)), and C-5 MTase from pea (Pradhan et al., *Nucleic Acids Res.*, 26(5):1214-1222 (1998)). Class II sequences have been detected in many species with a defining characteristic of the presence of an embedded chromodomain (Rose et al., *Nucleic Acids Res.*, 26(7):1628-1635 (1998)). The only full-length class II sequence is Cmt1 from Arabidopsis (Genbank #AF039364).

Class I enzymes are homologous to dnmt1 from mice (Bestor, T., et al., *EMBO J.*, 11(7):2611-2617 (1988)), the first cloned DNA methyltransferase. A knockout of dnmt1 in mice resulted in lethality during embryogenesis (Li et al., *Cell*, 69(6):915-926 (1992)). Dnmt1 has been used as a model for all class I enzymes though it has not been proven whether this is appropriate in plant systems. Antisense expression of MetI in Arabidopsis resulted in numerous developmental abnormalities (Finnegan et al., *Proc. Natl. Acad. Sci. U.S.A.*, 93(16):8449-8454 (1996)). Class I enzymes are thought to function as maintenance enzymes, though proteolytic cleavage could create de novo enzymes (Bestor, T.H., *EMBO J.*, 11(7):2611-2617

(1992)). CpG activity has been shown for dnmt1 in mice and humans. In peas it was found that pea C-5 MTase expressed in baculovirus displayed both CpG and CpNpG activity (Pradhan et al., *Nucleic Acids Res.*, 26(5):1214-1222 (1998)). In general, class I enzymes have a high level of expression in tissues that are actively dividing and are expressed at lower levels or silent in mature tissues.

There is little known regarding the function of class II enzymes. Cmt1 was detected as an Arabidopsis genomic sequence based on sequence homology to other methyltransferases. The C-terminal region contains the conserved methyltransferase domains and a chromodomain. The N-terminal region is much shorter than the N-terminal region of class I enzymes. Several commonly used ecotypes of Arabidopsis contain an allele of Cmt1 which is interrupted by a transposon insertion. These Cmt1 knockouts do not have any detectable phenotype. No other research has been published on the function of class II enzymes. Cmt1 is expressed only in floral tissues at very low levels. Degenerate PCR has been used to show the presence of Cmt1 homologs in a number of other plant species (Rose et al., *Nucleic Acids Res.*, 26(7):1628-1635 (1998)). In addition to finding homologs in other species, two sequences with similarity to Cmt1, Cmt2 and Cmt3, were identified in the Arabidopsis.

DNA methylation provides a mechanism for the mitotic propagation of epigenetic states. Epigenetic lineage-dependent patterns of gene expression have been studied the most in the germline and in somatic cell lineages in multicellular eukaryotes (Martienssen, R.A., et al., *Curr. Opin. Genet. and Develop.*, 5:234-242 (1995)). For example, in mice, the parentally imprinted genes *H19* and *Igf2r* are expressed in the embryo only when they are inherited via the female gamete. *Id.* In contrast, the *Igf2* gene is expressed only when inherited via the male gamete. *Id.* The human homologs of the *Igf2* and *H19* genes are linked and parentally imprinted as in the mouse. *Id.* Parental uniparental disomy for this chromosomal region (11p15) is associated with Beckwith-Wiedemann syndrome, which is believed to result from overexpression of *Igf2*. *Id.* In addition to overgrowth of certain organs, Beckwith-Wiedemann syndrome patients have a 700-fold predisposition to Wilms' tumor, and loss of heterozygosity in this region is found in many other tumors as well. *Id.* It has

also been shown that 60-70% of Wilms' tumor patients have biallelic expression of *Igf2, H19*, or both in tumor tissue, resulting from loss of imprinting rather than loss of heterozygosity. *Id.*

In plants, epigenetic changes in gene expression are considered to be easier to observe than in animals since there is little cell migration and clonal lineages stay together. *Id.* Moreover, because in plants the germline arises relatively late in development, many somatically variegated phenotypes can be followed into the next generation and are heritable to greater or lesser extents. *Id.* Parental imprinting of gene expression was first observed in plants at the *R* locus in maize. *Id.* Certain alleles condition a mottled phenotype in the alerone layer of the extra-embryonic endosperm when inherited paternally, but cause a fully colored phenotype when inherited maternally. *Id.* Genetic studies of modifier loci have revealed that it is the maternally inherited *R* allele that is imprinted to a high level of expression. *Id.* High levels of R expression correlate with demethylation of sites in the transcribed region in the maternally inherited allele. *Id.*

Plants transformed with additional copies of endogenous genes or with multiple copies of a foreign or exogenous gene (these endogenous and exogenous genes are often referred to as "transgenes") frequently display epigenetic inactivation. This phenomenon is known as "gene silencing" or "co-suppression". There are two types of "gene silencing" or "co-suppression". The first is "transcriptional silencing". In "transcriptional silencing", RNA production from the introduced transgene is repressed. The second type of "gene silencing" is "posttranscriptional silencing". In "posttranscriptional silencing", transcripts do not accumulate in the cytoplasm even though transcription rates are comparable with or are higher than those in cells where transcripts do accumulate.

Transcriptional silencing is associated with transgene methylation, particularly in the promoter (Finnegan, E.J., et al., *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49:223-47 (1998)). Posttranscriptional silencing, which affects both transgenes and homologous endogeneous genes, is also associated with transgene methylation, but within the coding sequence rather than the promoter. *Id.* It is believed that both

forms of gene silencing reflect normal, cellular defenses against invading or mobile DNAs. *Id.*

Currently, two classes of methyltransferase genes have been cloned in maize. The class I clone homolog is referred to as Zmet1 and the class II homolog Zmet2. The Zmet1 is a class I enzyme that was cloned by Paula Olhoft and Ron Phillips at the University of Minnesota. FIG. 4 is a summary of the major classes of 5-cytosine methyltransferases from mammals, *Arabidposis* and maize. The present invention herein relates to zmet2a and zmet2b methyltransferases.

SUMMARY OF THE INVENTION

In one embodiment, the present invention relates to an isolated and purified *Zea mays* zmet2a methyltransferase nucleic acid sequence. Specifically, the isolated and purified *Zea mays* zmet2a methyltransferase nucleic acid sequence of the present invention hybridizes to the nucleic acid sequences shown in FIG. 1A and 1B under stringent conditions. The zmet2a methyltransferase nucleic acid sequence encodes the enzyme zmet2a methyltransferase. The amino acid sequences for zmet2a methyltransferase is shown in FIG. 2A and FIG. 2B.

In another embodiment, the present invention further relates to recombinant expression cassettes comprising the isolated and purified zmet2a nucleic acid sequence described herein. Preferably, the recombinant expression cassettes further contain a promoter sequence and a polyadenylation signal sequence. The promoter sequence can be operably linked to the zmet2a nucleic acid sequence. The zmet2a nucleic acid sequence is operably linked to the polyadenylation signal sequence. Any promoter sequence can be used in the recombinant expression cassette, such as, but not limited to a constitutive or tissue specific promoter.

In another embodiment, the present invention also relates to a recombinant expression cassettes comprising one or more heterologous nucleic acid sequences. Such recombinant expression cassettes further contain a promoter sequence from the zmet2a nucleic acid sequence and a polyadenylation signal sequence. The promoter sequence is operably linked to the heterologous nucleic acid sequence. The

heterologous nucleic acid sequence is operably linked to the polyadenylation signal sequence. Any heterologous promoter sequence can be used in this recombinant expression cassette.

In a further embodiment, the present invention also relates to bacterial cells comprising at least one of the recombinant expression cassettes described herein. The bacterial cells can be *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*.

In a further embodiment, the present invention further relates to transgenic plant cells and transgenic plants containing the recombinant expression cassettes described herein. Monocotyledonous or dicotyledonous plant cells and plants can be transformed with the hereinbefore described recombinant expression cassettes. Plants which can be transformed with the recombinant expression cassettes of the present invention include, but are not limited to, *Zea mays*, *Oryza sativa*, *Secale cereale*, *Triticum aestivum*, *Daucus carota*, *Brassica oleracea*, *Cucumis melo*, *Cucumis sativus*, *Latuca sativa*, *Solanum tubersoum*, *Lycopersicon esculentum*, *Phaseolus vulgaris*, *Brassica napus*, etc. The present invention also relates to seed resulting from the transgenic plants of the present invention.

In a further embodiment, the present invention further provides methods of reducing or altering methyltransferase activity in a transgenic plant in order to increase transgene expression stability and/or to improve the yield or biochemical qualities of a plant as well as a method of silencing targeted genes in a plant *in vivo*. These methods comprise introducing into a plant a recombinant expression cassette comprising an appropriate plant promoter operably linked to a zmet2a methyltransferase nucleic acid sequence described herein in either the sense or antisense direction.

In a further embodiment, the present invention relates to an isolated and purified *Zea mays* zmet2b methyltransferase nucleic acid sequence. The zmet2b methyltransferase nucleic acid sequence of the present invention can be isolated using an isolated and purified partial *Zea mays* zmet2b methyltransferase nucleic acid sequence. The isolated and purified partial *Zea mays* zmeb2b methyltransferase

nucleic acid sequence can be used as a probe to isolate the zmet2b methyltransferase nucleic acid encoding zmet2b methyltransferase. Preferably, the isolated and purified partial *Zea mays* zmet2b methyltransferase nucleic acid described herein hybridizes to FIG. 23 under stringent conditions. The partial zmet2b methyltransferase nucleic acid sequence described herein encodes a portion of zmet2b methyltransferase. The partial amino acid sequence of zmet2b methyltransferase is shown in FIG. 24. The zmet2b methyltransferase nucleic acid sequence can be used in recombinant expression cassettes in the same manner as the isolated and purified zmet2a nucleic acid sequence described herein. Such recombinant expression cassettes can be used to create transgenic plants containing these recombinant expression cassettes. Additionally, the zmet2b methyltransferase nucleic acid sequence can be used to reduce or alter methyltransferase activity in transgenic plants in the same manner as the zmet2a methyltransferase nucleic acid sequence.

Definitions

Units, prefixes, and symbols can be denoted in the SI accepted form. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation, respectively. The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

As used herein, the term "plant" includes reference to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds and plant cells and progeny thereof. The class of plants which can be used in the methods of the present invention are generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants.

As used herein, "heterologous" when used to describe nucleic acids or polypeptides refers to nucleic acids or polypeptides that originate from a foreign species, or, if from the same species, are substantially modified from their original form. For example, a promoter operably linked to a heterologous structural gene is

from a species different from that from which the structural gene was derived, or, if from the same species, one or both are substantially modified from their original form.

A nucleic acid or polypeptide is "exogenous to" an individual plant is one which is introduced into the plant by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, and the like. Such a plant containing the exogenous nucleic acid is referred to herein as an R_1 generation transgenic plant. Transgenic plants which arise from sexual cross or by selfing are descendants of such a plant.

As used herein, "zmet2a methyltransferase gene" or "zmet2a methyltransferase nucleic acid" refers to a nucleic acid encoding zmet2a methyltransferase and which hybridizes under stringent conditions and/or has at least 60% sequence identity at the deduced amino acid level to the exemplified sequences provided herein. The zmet2a polypeptide encoded by the zmet2a methyltransferase gene has at least 55% or 60% sequence identity, typically at least 65% sequence identity, preferably at least 70% sequence identity, often at least 75% sequence identity, more preferably at least 80% sequence identity, and most preferably at least 90% sequence identity at the deduced amino acid level relative to the exemplary zmet2a methyltransferase sequences provided herein.

As used herein, "zmet2a methyltransferase nucleic acid" includes reference to a contiguous sequence from a zmet2a methyltransferase gene of at least 2454 nucleotides in length. In some embodiments the nucleic acid is preferably at least 2736 nucleotides in length (see FIG. 1A) and more preferably at least 2796 nucleotides in length (see FIG. 1B).

As used herein, "zmet2b methyltransferase gene" or "zmet2b methyltransferase nucleic acid" refers to a nucleic acid encoding zmet2b methyltransferase and which can be identified using the partial zmet2b methyltransferase nucleic acid shown in FIG. 23. The zmet2b methyltransferase gene

hybridizes under stringent conditions to the partial zmet2b methyltransferase nucleic acid shown in FIG. 23.

As used herein, "a partial zmet2b methyltransferase nucleic acid" includes reference to a contiguous sequence of at least 1181 nucleotides in length and which is from the zmet2b methyltransferase gene.

As used herein, "isolated" includes reference to material which is substantially or essentially free from components which normally accompany or interact with it as found in its naturally occurring environment. The isolated material optionally comprises material not found with the material in its natural environment.

As used herein, "nucleic acid" includes reference to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

As used herein, "operably linked" includes reference to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and, where necessary to joint two protein coding regions, contiguous and in the same reading frame.

In the expression of transgenes, one of ordinary skill in the art will recognize that the inserted nucleic acid sequence need not be identical and may be "substantially identical" to a sequence of the gene from which it was derived. As explained below, these variants are specifically covered by this term.

In the case where the inserted nucleic acid sequence is transcribed and translated to produce a functional zmet2a and/or zmet2b methyltransferase

polypeptide, one of ordinary skill in the art will recognize that because of codon degeneracy, a number of nucleic acid sequences will encode the same polypeptide. These variants are specifically covered by the term "zmet2a methyltransferase nucleic acid sequence" or "zmet2b methyltransferase nucleic acid sequence". In addition, the term specifically includes those full length sequences substantially identical (determined as described below) with a zmet2a and/or zmet 2b methyltransferase gene sequence which encode proteins that retain the function of the zmet2a and/or zmet2b methyltransferase. Thus, in the case of the zmet2a and/or zmet2b methyltransferase genes described herein, the term includes variant nucleic acid sequences which have substantial identity with the sequences disclosed herein and which encode proteins capable of reducing or regulating DNA methylation in a transgenic plant for various purposes as well as silencing target genes in a plant using the nucleic acid sequences described herein.

Two nucleic acids or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The term "complementary to" is used herein to mean that the complementary sequence is identical to all or a specified contiguous portion of a reference nucleic acid sequence. Sequence comparisons between two (or more) nucleic acids or polypeptides are typically performed by comparing sequences of two optimally aligned sequences over a segment or "comparison window" to identify and compare local regions of sequence similarity. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman, *Ad. App. Math.* 2: 482 (1981), by the homology alignment algorithm of Neddleman and Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman, *Proc. Natl. Acad. Sci.* (U.S.A.) 85:2444 (1988), by computerized implementation of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (hereinafter "GCG"), 575 Science Dr., Madison, WI), or by inspection.

"Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, where the portion of the nucleic acid

sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

The term "substantial identity" of nucleic acid sequences means that a nucleic acid comprises a sequence that has at least 55% or 60% sequence identity, generally at least 65%, preferably at least 70%, often at least 75%, more preferably at least 80% and most preferably at least 90%, compared to a reference sequence using the programs described above (preferably BESTFIT) using standard parameters. One of ordinary skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid sequences for those purposes normally means sequence identity of at least 55% or 60%, preferably at least 70%, more preferably at least 80%, and most preferably at least 95%. Polypeptides having "sequence similarity" share sequences as noted above except that residue positions which are not identical may differ by conservative amino acid changes. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

Another indication that nucleic acid sequences are substantially identical is if two molecules hybridize to each other under appropriate conditions. Appropriate

conditions can be high or low stringency and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C to about 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH 0) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent wash conditions are those in which the salt concentration is about 0.22 molar at pH 7 and the temperature is at least about 50°C. However, nucleic acids which do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Nucleic acids of the present invention can be identified from a cDNA or genomic library prepared according to standard procedures and the nucleic acids disclosed here used as a probe. For example, stringent hybridization conditions will typically include at least one low stringency wash using 0.3 molar salt (e.g., 2X SSC) at 65°C. The washes are preferably followed by one or more subsequent washes using 0.03 molar salt (e.g., 0.2X SSC) at 50°C, usually 60°C, or more usually 65°C. Nucleic acid probes used to isolate the nucleic acids are preferably at least 100 nucleotides in length.

As used herein, a homologue of a particular zmet2a and/or zmet2b methyltransferase gene is a second gene (either in the same species or in a different species) which encodes a protein having an amino acid sequence having at least 50% identity or 75% similarity to (determined as described above) to a polypeptide sequence in the first gene product.

As used herein, "nucleotide binding site" or "nucleotide binding domain" includes reference to a region consisting of kinase-1a, kinase 2, and kinase 3a motifs, which participates in ATP/GTP-binding. Such motifs are described for instance in Yu *et al.*, *Proc. Acad. Sci. USA* 93:11751-11756 (1996); Mindrinos, *et al.*, *Cell* 78:1089-1099 and Shen *et al.*, FEBS, 335:380-385 (1993).

As used herein, "tissue-specific promoter" includes reference to a promoter in which expression of an operably linked gene is limited to a particular tissue or tissues.

As used herein "recombinant" includes reference to a cell, or nucleic acid, or vector, that has been modified by the introduction of a heterologous nucleic acid or the alteration of a native nucleic acid to a form not native to that cell, or that the cell is derived from a cell so modified. For example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

As used herein, a "recombinant expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements which permit transcription of a particular nucleic acid in a target cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the recombinant expression cassette portion of the expression vector includes a nucleic acid to be transcribed, and a promoter.

As used herein, "transgenic plant" includes reference to a plant modified by introduction of a heterologous nucleic acid. Generally, the heterologous nucleic acid is a zmet2a and/or zmet2b methyltransferase structural or regulatory gene or subsequences or combinations thereof.

As used herein, "hybridization complex" includes reference to a duplex nucleic acid sequence formed by selective hybridization of two single-stranded nucleic acids with each other.

As used herein, "amplified" includes reference to an increase in the molarity of a specified sequence. Amplification methods include the polymerase chain reaction (hereinafter "PCR"), the ligase chain reaction (hereinafter "LCR"), the transcription-based amplification system (hereinafter "TAS"), the self-sustained sequence replication system (hereinafter "SSR"). A wide variety of cloning methods,

host cells, and *in vitro* amplification methodologies are well-known to persons of ordinary skill in the art.

As used herein, "nucleic acid sample" includes reference to a specimen suspected of comprising a zmet2a and/or zmet2b methyltransferase gene.

SEQUENCE LISTINGS

The present application contains a number of nucleotide sequences and amino acid sequences. For the nucleotide sequences, the base pairs are represented by the following base codes:

Symbol	Meaning
A	A; adenine
C	C; cytosine
G	G; guanine
T	T; thymine
U	U; uracil
M	A or C
R	A or G
W	A or T/U
S	C or G

Symbol	Meaning
Y	C or T/U
K	G or T/U
V	A or C or G; not T/U
H	A or C or T/U; not G
D	A or G or T/U; not C
B	C or G or T/U; not A
N	(A or C or G or T/U)

The amino acids shown in the application are in the L-form and are represented by the following amino acid-three letter abbreviations:

Abbreviation	Amino acid name
Ala	L-Alanine
Arg	L-Arginine
Asn	L-Asparagine
Asp	L-Aspartic Acid
Asx	L-Aspartic Acid or Asparagine
Cys	L-Cysteine
Glu	L-Glutamic Acid

Gln	L-Glutamine
Glx	L-Glutamine or Glutamic Acid
Gly	L-Glycine
His	L-Histidine
Ile	L-Isoleucine
Leu	L-Leucine
Lys	L-Lysine
Met	L-Methionine
Phe	L-Phenylalanine
Pro	L-Proline
Ser	L-Serine
Thr	L-Threonine
Trp	L-Tryptophan
Tyr	L-Tyrosine
Val	L-Valine
Xaa	L-Unknown or other

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A shows the nucleic acid sequence of the zmet2a methyltransferase gene containing 2736 basepairs. FIG. 1B shows the nucleic acid sequence of the zmet2a methyltransferase gene containing 2796 basepairs.

FIG. 2A shows the amino acid sequence of the zmet2a methyltransferase containing 912 amino acids and which is encoded by the nucleic acid sequence shown in FIG. 1A. FIG. 2B shows the amino acid sequence of the zmet2a methyltransferase containing 932 amino acids and which is encoded by the nucleic acid sequence shown in FIG. 1B.

FIG. 3 shows the PCR primers used to sequence the zmet2a methyltransferase gene.

FIG. 4 is a summary of the major classes of 5-cytosine methyltransferases from mammals, *Arabidopsis* and maize.

FIG. 5 shows the genomic sequence of zmet2a methyltransferase gene and the retrotransposon SPRITE-1, along with intron-exon divisions, a restriction site map and a primer map.

FIG. 6 lists the World Wide Web sites used to process the sequence data for the zmet2a methyltransferase gene.

FIG. 7 shows a Southern blot of B73 DNA digested with *Hind*III and probed with clone CGET064. The Southern blot shows the presence of multiple copies of zmet2a or zmet2a-like genes in the B73 genome. DNA from B73 was digested with HindIII and probed with clone CGET064 which does not contain a HindIII site. The gene cloned and sequenced is represented by the upper band.

FIG. 8 shows the alignment of the amino acid sequence from zmet2a with the amino acid sequence of *Arbadiopsis* chromosmethylase *CMT1* (AF039367) and the C-terminal methylase domains from the DNA methyltransferases of maize zmet1 (AF063403) and *Arabidopsis MET1* (P34881). Zmet2a shows similarity along the entire length of *CMT1* but significant similarity with zmet1 and Met1 exists only in the conserved motifs. Bold, uppercase, normal uppercase letters, and lower case letters indicate identity, conservation, and differences in amino acid sequences relative to zmet2a respectively. Dashes in the sequences are gaps introduced by CLUSTAL W to optimize the alignments. The location of the six conserved methylase motifs are indicated in the sequence. The chromodomain is located upstream and adjacent to motif IV. The *Mu* insertion into the coding region of motif IX alters zmet2a function resulting in decreased methylation at CpNpG sites. Putative nuclear loalization signal peptides, NLS (N. Raikhel, *Plant Physiol.* 100, 1627 (1992)) are positioned in the N-terminal portion of the protein.

FIG. 9 lists the putative identification of zmet2a amino acids involved in catalysis by comparison with amino acids of M.Hhal with known catalytic functions. The amino acids of M.Hhal with catalytic functions were determined by crystalography by Cheng et al., *Cell*, 74:299-307 (1993). Amino acid of zmet2a are numbered as in Figure 7.

FIG. 10 shows southern analysis of repetitive DNA methylation patterns. Total genomic DNA (5 μg per lane) from an F_4 derived F_5 family segregating for

zmet2a:*Mul* was digested with isoschizomers *Hpa*II and *Msp*I which recoginze the sequence CCGG. Digested DNA ws electophoresed through 0.8% agarose, transferred to nylon membrane, and hybridized with probes for repetitive DNA; the 9kb 26s-5.8s-17s ribosomal repeat (FIG. 10A), 5s ribosomal repeat (FIG. 10B), and a centromeric repeat pSau3a9 (FIG. 10C). Decreased methylation is observed in mutant plants (- -) relative to nonmutant plants (+ +) digested with *Msp*I which is sensitive to methylation at ^{me}CpCpG sequnces. No changes in methylation patterns at ^{me}CpG sits are observed in mutant plants as indicated by the lack of digestion with *Hpa*II. Plants heterozygous for zmet2a:*Mul* (+ -) also show decreases at ^{me}CpCpg sites.

FIG. 11 shows gels from a Southern analysis which demonstrate that plants homozygous for zmet2a::Mu1 have decreased methylation at CpNpG sites. More sites cut with restriction enzymes that are sensitive to methylation at CpNpG sites in zmet2a:*Mul* plants. *Eco*RII is sensitive to methylation at CC*A/TGG sites where * indicates the senitive cytosine (FIG. 11A). *Bgl*II is sensitive to methylation at AGATC*T sites (FIG. 11B). *Pst*I is sensitive to methylation at C*TGCAG sites (FIG. 11C). *Bam*HI is sensitive to methylation at GGATC*C sites (FIG. 11D). *Ava*II is sensitive to methylation at GGA/TC*C sites (FIG. 11E). Changes at CpG sites cannot be sparated from CpCpG in the *Ava*II digests. DNA from the same plants as those in Figure 10 were digested and hybridized with the repetitive probes as described herein.

FIG. 12 shows the cytosine methylation levels in an F4 derived F5 segregating line for zmet2a::Mu1. 5-methylcytosine content of DNA extracted from tissue of immature 5th –7th leaves was determined by reverse phase HPLC using the method of Gehrike et al. Values were obtained from three wildtype plants, seven heterozygous plants and five homozygous plants. Two samples were run for each plant. Percentages of 5mC content [5mC/(5mC + C)] were calculated from concentrations determined from integration of peak and comparison to known standards.

FIG. 13 shows gels from a Southern analysis which demonstrate that plants homozygous for zmet2a::Mu1 having a reduced level of methylation that is stable

over generations. Two F_2 derived F_3 families homozygous for zmet2a:Mul, B5 and B6, were self pollinated to the F_6 generation. Two lineages from B5 and three lineages from B4 were grown at the University of Wisconsin, West Madison Agronomy Farm in 1999. Methylation levels are consistent across generations. Once zmet2a:*Mul* is in a homozygous state, methylation is reduced to a specific level and no further reductions occur. Dilution of methylation is not observed in each successive generation. DNA from leaf tissue was digested with *Msp*I and the Southern blot was hybridized with 9kb ribosomal repetitive probe.

FIG. 14 shows gels from a Southern analysis which demonstrate that methylation levels are restored to nonmutant parental levels in backcross progeny homozygous for wildtype zmet2a. An F1 hybrid of an F4 line homozygous for zmet2a::*Mul* (lanes 1-3) and the inbred line Mo17 (lanes 4-6) was backcrossed to the nonmutant Mo17 parent ot generate plants homozygous wildtype and plants heterozygous for zmet2a:*Mul*. F1 plants (lanes 7-11) have methylation levels intermediate those of the parents. BC1 progeny heterozygous for zmet2a:*Mul* (lanes 12-17) have methylation levels similar to the F1. BC1 plants restored to wild-type zmet2a (lanes 18-21) have remethylation to levels comparable to the nonmutant parent line. Complete or near complete remethylation has occurred within one sexual generation. DNA was extracted from the 4^{th} – 6^{th} immature leaves of greenhouse grown seedlings, digested with *Pst*I which is sensitive to methylation at ^{me}CTGCAG sequences, and hybridized to the pSau3a9 centromeric repeat.

FIG. 15 shows gels from a Southern analysis which demonstrate the expression of zmet2a in different tissues during development. Southern blots were produced with cDNA's synthesized from mRNA extracted from embryos 24 days after pollination (hereinafter "DAP"), young leaves, immature ear, immature tassel, BMS callus, and 10 day old seedlings. Figure 15A shows the ethidium bromide stained gel. All lanes were loaded with 750 ng of cDNA except for the 10 day seedlings, of which 280 ng was loaded due to the limited amount available. The cDNA's were quantified by spectrophotometry. The marker lane contains 800 ng of lambda DNA digested with *Hind*III. Figure 15B shows the Southern blot hybridized with the zmet2a cDNA probe. Hybridization is observed in tissues that are actively

undergoing cell division. Figure 15C shows the same blot hybridized to a ubiquitin probe to show cDNA loading variation.

FIG. 16 shows the structure of maize retrotransposon SPRITE-1 and sequence of Long Terminal Repeat (hereinafter "LTR") components. FIG. 16A shows that SPRITE-1 consists of long terminal direct repeats, a tRNA primer binding site (hereinafter "PBS"), coding sequence for proteins necessary for replication and transposition, and a polypurine tract (hereinafter "PPT"). FIG. 16B identifies the sequences for the 5' and 3' LTR , PBS and PPT. Each LTR has a 3 base pair inverted repeat which is also shown in the drawing. A putative TATA box is underlined and the putative transcription start site is italicized. The 5 base pair host insertion site duplications are also identified.

FIG. 17 shows the alignments of the conserved protein motifs of the Ty1/copia elements with SPRITE-1. The maize retrotransposon SPRITE-1 is aligned with the retrotransposon hopscotch (U2626) from maizze, retrofit (U72725) from rice, an unpublished *Arabidopsis* retrotransposon (AC006528) and the copia element from Drosophila (M11240).

FIG. 18 shows that the SPRITE-1 copy number and insertion sites differ among maize inbred lines. DNA (7 μg) from inbred maize lines, barley, ice, rye, wheat, and potato was digested with BcoRI which does not cut within the retroelement sequence. The Southern blot was hybridized with a 950 bp SPRITE-1 fragment which includes the 5' untranslated sequence and 5' sequence putatively coding for the *gag* protein but does not include the conserved *gag* motif or the 5' terminal repeat.

FIG. 19 shows the identification of inbred lines containing a SPRITE-1 insertion in zmet2a. PCR was conducted on maize inbred lines from various origins using a primer upstream of the SPRITE-1 insertion site 15F in conjuction with a SPRITE-1 specific primer 18R or a zmet2a primer downstream of the element 8R. The upper panel (15F/18R) show the inbreds that do not have a SPRITE-1 insertion. The lower panel (15F/18R) shows that Mol17 and A682 have a SPRITE-1 insertion

into zmet2a. A682 has an amplification product from both primer sets indicating that it may be hemizygous for SPRITE-1.

FIG. 20 shows expression of retroelement SPRITE-1. Figure 20A shows a Southern blot of cDNAs from roots, immature embryo 24 days after pollination (hereinafter, "DAP"), young leaf, young leaf with inacive zmet2a immature ear, immature tassel, mature pollen, Black Mexican Sweet (hereinafter, "BMS") callus, and 10 day seedling, hubridized with a SPRITE-1 probe. Transcription of SPRITE-1 is evident as indicated by the hybridization to cDNA from embryo, and leaf tissue. Expression is highest in leaf tissue with significantly more expression being observed in leaf tissue from zmet2a:Mul plants that have decreased CpNpG methylation. FIG. 20B shows the same Southern blot hybridized to a ubiquitin probe as a loading control.

FIG. 21 shows that the presence of a SPRITE-1 insertion into a zmet2a intron does not alter transcript splicing. Fragmetns spanning the SPRITE-1 insertion and downstream from the insertion site were amplified by PCR from cDNA's. FIG. 21A shows a scaled representation of zmet2a. Exons are represented by large blocks while the intervening introns are depicted by lines. The insertion of the retroelement is indicated above the zmet2a diagram. The element is inserted in the opposite orientation relative to zmet2a as indicated by the boxed arrows which represent the direct repeats. Positions of the primers used to generate fragments are indicated below the zmet2a diagram. Fragments were amplified from B73 (FIG. 21B) immature ear cDNA which does not contain the retroelement insertion and Mo17 (M) embryo 24 days after pollination cDNA (FIG. 21B) and Mo17 (M) 10 day seedling cDNA (FIG. 21C). No differences were observed on the ethidium bromide stained gel of the PCR products. FIGS. 21B and 21C show hybridization of a near full length B73 cDNA probe to a Southern blot of the PCR fragments.

FIG. 22 shows the methylation status of SPRITE-1. DNA from immature leaves was digested with methylation sensitive restriction enzymes. Southern blots were hybridized with a 970 base pair fragment from the 5' end of the untranslated region of SPRITE-1. There are 5 BstNI/EcoRII sites, 1 MspI/HpaII sites and 1 PstI

site within the sequence context of this probe. Nearly all sites are methylated in this region.

FIG. 23 shows a partial nucleic acid sequence of the zmet2b methyltransferase gene.

FIG. 24 shows a partial amino acid sequence of the zmet2b methyltransferase encoded by the partial nucleic acid sequence shown in FIG. 23.

FIG. 25 shows a comparison of a portion of the amino acid sequence for zmet2a methyltransferase with a portion of the amino acid sequence for zmet2b methyltransferase.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In one embodiment, the present invention relates to a zmet2a methyltransferase gene. The zmet2a methyltransferase gene of the present invention encodes a class II methyltransferase gene which controls CpNpG methylation. Nucleic acid sequences from the zmet2a methyltransferase gene described herein can be used to reduce or to alter the level of DNA methylation in a plant. In addition, the zmet2a nucleic acid sequence described herein can be used to methylate a targeted gene in a plant *in vivo* to "silence" or "knock-out" said gene.

In another embodiment, the present invention relates a zmet2b methyltransferase gene. The zmet2b methyltransferase gene can be isolated using a partial zmet2b methyltransferase gene described herein. Like the zmet2a methyltransferase gene, the zmet2b methyltransferase gene encodes a class II methyltransferase gene which controls CpNpG methylation. Nucleic acid sequences encoding the zmet2b methyltransferase gene can be used in the same manner as the nucleic acid sequence encoding the zmet2a methyltransferase gene to reduce or to alter the level of DNA methylation in a plant. In addition, the zmet2b nucleic acid sequence can be used to methylate a targeted gene in a plant *in vivo* to "silence" or "knock-out" said gene.

The present invention is applicable to a broad range of types of monocotyledonous and dicotyledonous plants, including, but not limited to, *Zea mays*, *Oryza sativa*, *Secale cereale*, *Triticum aestivum*, *Daucus carota*, *Brassica oleracea*, *Cucumis melo*, *Cucumis sativus*, *Latuca sativa*, *Solanum tubersoum*, *Lycopersicon esculentum*, *Phaseolus vulgaris*, and *Brassica napus*.

The nucleic acids of the present invention can be used in marker-aided selection. Marker-aided selection does not require the complete sequence of the gene or precise knowledge of which sequence confers which specificity. Instead, partial sequences can be used as hybridization probes or as the basis for oligonucleotide primers to amplify by PCR or other methods to follow the segregation of chromosome segments containing the zmet2a and/or zmet2b methyltransferase gene(s) in plants. Because the zmet2a or zmet2b methyltransferase marker is the gene itself, there can be negligible recombination between the marker and the methylated phenotype. Thus, the nucleic acids of the present invention can be used to provide an optimal means to DNA fingerprint class II DNA methyltransferases in other cultivars and wild germplasm. This can be used to indicate if other germplasm accessions and cultivars carry the same zmet2a and/or zmet2b methyltransferase genes.

Preparation of the Nucleic acids of the Present Invention

Generally, the nomenclature and the laboratory procedures involved with recombinant DNA technology described below are those well known and commonly employed by those of ordinary skill in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally, enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. These techniques and various other techniques are generally performed according to Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989).

The isolation of zmet2a and/or zmet2b methyltransferase gene(s) can be accomplished via a number of techniques. For instance, oligonucleotide probes based on the sequences disclosed herein can be used to identify the desired gene in a cDNA

or genomic DNA library. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a cDNA library, mRNA is isolated from the desired organ of a particular plant, such as shoots from *Zea mays*, and a cDNA library which contains the zmet2a or zmet2b methyltransferase gene transcript is prepared from the mRNA. Alternatively, cDNA may be prepared from mRNA extracted from other tissues in which the zmet2a or zmet2b methyltransferase gene or homologs are expressed.

The cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned zmet2a and/or zmet2b methyltransferase gene or partial sequence from either thereof (such as the partial zmet2b methyltransferase nucleic acid sequence shown in FIG. 23). Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species.

Those of ordinary skill in the art will appreciate that various degrees of stringency of hybridization can be employed in the assay; and either the hybridization or the wash medium can be stringent. As the conditions for hybridization become more stringent, there is a greater degree of complementarity required between the probe and the target for duplex formation to occur. The degree of stringency can be controlled by temperature, ionic strength, pH and the presence of a partially denaturing solvent such as formamide. For example, the stringency of hybridization is conveniently varied by changing the polarity of the reactant solution through manipulation of the concentration of formamide within the range of 0% to 50%.

Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (hereinafter "PCR") technology can be used to amplify the sequences of the zmet2a and/or zmet2b methyltransferase and related genes directly from genomic DNA, from cDNA, from genomic libraries or from cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid

sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes.

The degree of complementarity (sequence identity) required for detectable binding will vary in accordance with the stringency of the hybridization medium and/or wash medium. The degree of complementarity will optimally be 100 percent; however, it should be understood that minor sequence variations in the probes and primers may be compensated for by reducing the stringency of the hybridization and/or wash medium as described earlier.

Appropriate primers and probes for identifying zmet2a and/or zmet2b methyltransferase nucleic acid sequences from plant tissues are generated from a comparison of the sequences provided herein. For a general overview of PCR see *PCR Protocols: A Guide to Methods and Applications*. (Innis, M, Gelfand, D., Snisky, J. and White, T., eds), *Academic Press*, San Diego (1990), incorporated herein by reference.

Nucleic acids may also be synthesized by well-known techniques as described in the technical literature. See e.g., Curruthers *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and Adams *et al.*, *J. Am. Chem. Soc.* 105:661 (1983). Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Proteins of the Present Invention

The present invention further provides for isolated zmet2a and/or zmet2b methyltransferases encoded by the zmet2a and/or zmet2b methyltransferase nucleic acids disclosed herein. One of ordinary skill in the art will recognize that nucleic acids encoding a functional zmet2a or zmet2b methyltransferase need not have a sequence identical to the exemplified genes disclosed herein. For example, because of codon degeneracy, a large number of nucleic acid sequences can encode the same

polypeptide. In addition, the polypeptides encoded by the zmet2a and/or zmet2b methyltransferase genes, like other proteins, have different domains which perform different functions. Specifically, zmet2a methyltransferase has ten (10) domains. These ten domains are identified as follows: I, chromodomain β2, chromodomain β3, IV, VI, VIII, IX and X. The ten domains and their sequence ranges (as shown in SEQ ID NO:2) are listed below in Table 1:

TABLE 1

Domain	Amino Acid Sequence Range
I	244-271
Chromodomain β2	366-379
Chromodomain β3	380-388
IV	411-434
VI	456-476
VIII	496-520
IX	723-746
X	751-775

Domains I and X are involved in binding AdoMet, which is source of the methyl group to be transferred during DNA methylation. Domain IV contains a catalytic domain. Domain VI aids in the positioning of domain IV. Domain VIII aids in DNA binding by neutralizing the charge of the phosphodiester backbone. The region between domain VIII and domain IX defines the sequence specificity of the zmet2a methyltransferase enzyme. Thus, the zmet2a methyltransferase gene sequences need not be full length, so long as the desired functional domain of the protein is expressed.

The zmet2a methyltransferase protein is at least 912 amino acid residues in length (see FIG. 2A), preferably, 932 amino acid residues in length (see FIG. 2B). However, those of ordinary skill in the art will appreciate that amino acid deletions, substitutions, or additions to the zmet2a methyltransferase protein will typically yield a enzyme possessing methylating characteristics similar or identical to that of the full length sequence. Thus, full length zmet2a methyltransferase proteins modified by 1,

2, 3, 4, or 5 deletions, substitutions, or additions, generally provide an effective degree of methylation relative to the full-length protein.

A partial amino acid sequence of the zmet2b methyltransferase protein is provided for in FIG. 24 and is 256 amino acids in length.

Modified protein chains can also be readily designed utilizing various recombinant DNA techniques well known to those of ordinary skill in the art. For example, the chains can vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. Modification can also include swapping domains from the proteins of the present invention with related domains from other class II methyltransferases.

The present invention also provides antibodies which specifically react with the zmet2a and/or zmet2b methyltransferase(s) of the present invention under immunologically reactive conditions. An antibody immunologically reactive with a particular antigen can be generated *in vivo* or by recombinant methods such as by selection of libraries of recombinant antibodies in phage or similar vectors. The term "immunologically reactive conditions" as used herein, includes reference to conditions which allow an antibody, generated to a particular epitope of an antigen, to bind to that epitope to a detectably greater degree than the antibody binds to substantially all other epitopes, generally at least two times above background binding, preferably at least five times above background. Immunologically reactive conditions are dependent upon the format of the antibody binding reaction and typically are those utilized in immunoassay protocols.

The term "antibody" as used herein, includes reference to an immunoglobulin molecule obtained by *in vitro* or *vivo* generation of the humoral response, and includes both polyclonal and monoclonal antibodies. The term also includes genetically engineered forms such as chimeric antibodies (e.g., humanized murine antibodies), heteroconjugate antibodies (e.g., bispecific antibodies), and recombinant single chain Fv fragments (scFv). The term "antibody" also includes antigen binding forms of antibodies (e.g., Fab', $F(ab')_2$, Fab, Fv, and, inverted IgG. See, Pierce

Catalog and Handbook, 1994-1995) Pierce Chemical Co., Rockford, IL). An antibody immunologically reactive with a particular antigen can be generated *in vivo* or by recombinant methods such as selection of libraries of recombinant antibodies in phage or similar vectors (*See, e.g.* Huse *et al.*, (1989) *Science* 246:1275-1281; and Ward, *et al.*, (1989) *Nature* 341:544-546; *and Vaughan* et al., (1996) *Nature Biotechnology*, 14:309-314).

Many methods of making antibodies are known to persons of ordinary skill in the art. A number of immunogens are used to produce antibodies specifically reactive to the zmet2a and/or zmet2b methyltransferase(s) of the present invention under immunologically reactive conditions. An isolated recombinant, synthetic, or native zmet2a and/or zmet2b methyltransferase(s) of the present invention is the preferred immunogens (antigen) for the production of monoclonal or polyclonal antibodies.

The zmet2a and/or zmet2b methyltransferase(s) is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies can be generated for subsequent use in immunoassays to measure the presence and quantity of the zmet2a and/or zmet2b methyltransferases. Methods of producing monoclonal or polyclonal antibodies are known to those of skill in the art (*See,* Coligan (1991) *Current Protocols in Immunology* Wiley/Greene, NY; and Harlow and Lane (1989) *Antibodies: A Laboratory Manual* Cold Spring Harbor Press, NY); Goding (1986) *Monoclonal Antibodies: Principles and Practice* (2d ed.) Academic Press, New York, NY).

Frequently, the zmet2a and/or zmet2b methyltransferase(s) and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionucleotides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, chemiluminescent moieties, magnetic particles, and the like. Patents teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

The antibodies of the present invention can be used to screen plants for the expression of the zmet2a and/or zmet2b methyltransferase(s). The antibodies of the present invention are also used for affinity chromatography in isolating zmet2a and/zmet2b methyltransferase(s).

The present invention further provides zmet2a and/or zmet2b methyltransferase polypeptides that specifically bind, under immunologically reactive conditions, to an antibody generated against a defined immunogen, such as an immunogen consisting of the polypeptides of the present invention. For example, immunogens will generally be at least 912 contiguous amino acids from the zmet2a methyltransferase polypeptide of the present invention. Nucleic acids which encode such cross-reactive zmet2a and/or zmet2b methyltransferase polypeptides are also provided by the present invention. The zmet2a/zmet2b methyltransferase polypeptides can be isolated from any number of plants as discussed earlier. Preferred plants are *Zea mays*, *Oryza* sativa, *Secale cereale*, *Triticum aestivum*, *Daucus carota*, *Brassica oleracea*, *Cucumis melo*, *Cucumis sativus*, *Latuca sativa*, *Solanum tubersoum*, *Lycopersicon esculentum*, *Phaseolus vulgaris*, and *Brassica napus*.

As used herein, the term, "specifically binds" includes reference to the preferential association of a ligand, in whole or part, with a particular target molecule (i.e., "binding partner" or "binding moiety" relative to compositions lacking that target molecule). It is, of course, recognized that a certain degree of non-specific interaction may occur between a ligand and a non-target molecule. Nevertheless, specific binding, may be distinguished as mediated through specific recognition of the target molecule. Typically, specific binding results in a much stronger association between the ligand and the target molecule than between the ligand and non-target molecule. Specific binding by an antibody to a protein under such conditions requires an antibody that is selected for its specificity for a particular protein. The affinity constant of the antibody binding site for its cognate monovalent antigen is at least 10^7, usually at least 10^9, more preferably at least 10^{10}, and most preferably at least 10^{11} liters/mole. A variety of immunoassay formats are appropriate for selecting antibodies specifically reactive with a particular protein. For example, solid-phase

ELISA immunoassays are routinely used to select monoclonal antibodies specifically reactive with a protein (See Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific reactivity). The antibody may be polyclonal but preferably is monoclonal. Generally, antibodies cross-reactive to zmet2a and/or zmet2b methyltransferases are removed by immunoabsorbtion.

Immunoassays in the competitive binding format are typically used for cross-reactivity determinations. For example, an immunogenic zmet2a and/or zmet2b methyltransferase polypeptide is immobilized to a solid support. Polypeptides added to the assay compete with the binding of the antisera to the immobilized antigen. The ability of the above polypeptides to compete with the binding of the antisera to the immobilized zmet2a and/zmet2b methyltransferase polypeptides are compared to the immunogenic zmet2a and/or zmet2b methyltransferase polypeptide(s). The percent cross-reactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% cross-reactivity with such proteins as zmet2a and/or zmet2b methyltransferase(s) are selected and pooled. The cross-reacting antibodies are then removed from the pooled antisera by immunoabsorbtion with the non-zmet2a and/or non-zmet2b methyltransferase polypeptide(s).

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay to compare a second "target" polypeptide to the immunogenic polypeptide. In order to make this comparison, the two polypeptides are each assayed at a wide range of concentrations and the amount of each polypeptide required to inhibit 50% of the binding of the antisera to the immobilized protein is determined using standard techniques. If the amount of the target polypeptide required is less than twice the amount of the immunogenic polypeptide that is required, then the target polypeptide is said to specifically bind to an antibody generated to the immunogenic protein. As a final determination of specificity, the pooled antisera is fully immunoabsorbed with the immunogenic polypeptide until no binding to the polypeptide used in the immunoabsorbtion is detectable. The fully immunoabsorbed antisera is then tested for reactivity with the test polypeptide. If no reactivity is

observed, then the test polypeptide is specifically bound by the antisera elicited by the immunogenic protein.

Production of Recombinant Expression Cassettes

Isolated sequences prepared as described herein can then be used to provide recombinant expression cassettes. One of ordinary skill in the art will recognize that the nucleic acids used in the recombinant expression cassettes described herein encoding a functional zmet2a and/or zmet2b methyltransferase(s) need not have a sequence identical to the exemplified genes disclosed herein. In addition, the polypeptides encoded by the zmet2a and/or zmet2b methyltransferase genes, like other proteins, have different domains which perform different functions. Thus, the zmet2a and/or zmet2b methyltransferase gene sequences need not be full length, so long as the desired functional domain of the protein is expressed.

A DNA sequence coding for the desired zmet2a and/or zmet2b methyltransferase polypeptide(s), for example a cDNA or a genomic sequence encoding a full length protein, can be used to construct a recombinant expression cassette which can be introduced into a desired plant. An expression cassette will typically comprise the zmet2a and/or zmet2b methyltransferase nucleic acid(s) operably linked in either the sense or antisense direction to transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the zmet2a and/or zmet2b methyltransferase gene(s) in the intended tissues for the transformed plant.

For example, a plant promoter fragment may be employed which will direct expression of the zmet2a and/or zmet2b methyltransferase in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters includes the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1' or 2' - promoter derived from T-DNA of *Agrobacterium tumefaciens*, and ubiquitin other transcription initiation regions from various plant genes known to those of ordinary skill in the art.

Alternatively, the plant promoter may direct expression of the zmet2a and/or zmet2b methyltransferase gene in a specific tissue or may be otherwise under more precise environmental or developmental control. Such promoters are referred to here as "inducible" promoters. Examples of environmental conditions that may effect transcription by inducible promoters include pathogen attack, anaerobic conditions, or the presence of light.

Examples of promoters under developmental control include promoters that initiate transcription only in certain tissues, such as leaves, roots, fruit, seeds, or flowers. The operation of a promoter may also vary depending on its location in the genome. Thus, an inducible promoter may be fully or partially constitutive in certain locations.

The endogenous promoters from the zmet2a and/or zmet2b methyltransferase genes of the present invention can be used to direct expression of the genes. These promoters can also be used to direct expression of heterologous structural genes. The promoters can be used, for example, in recombinant expression cassettes to drive expression of genes to produce DNA methyltransferase in a particular cell or tissue.

To identify the promoters, the 5' portions of the clones described herein are analyzed for sequences characteristic of promoter sequences. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually 20 to 30 base pairs upstream of the transcription start site. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) N G. J. Messing et al., in *Genetic Engineering in Plants*, pp. 221-227 (Kosage, Meredith and Hollaender, eds. 1983).

If proper polypeptide expression is desired, a polyadenylation region at the 3'-end of the zmet2a or zmet2b methyltransferase coding region should be included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences from the zmet2a and/or zmet2b methyltransferase gene(s) will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulforon.

As discussed above, the zmet2a and/or zmet2b methyltransferase gene(s) can be inserted into a recombinant expression cassette in the antisense direction. Expression of the zmet2a and/or zmet2b methyltransferase gene(s) in antisense direction will result in the production of antisense RNA. As is well known, a cell manufactures protein by transcribing the DNA of the gene encoding a protein to produce RNA, which is then processed to messenger RNA (mRNA) (e.g., by the removal of introns) and finally translated by ribosomes into protein. This process may be inhibited in the cell by the presence of antisense RNA. The term antisense RNA means an RNA sequence which is complementary to a sequence of bases in the mRNA in question in the sense that each base (or the majority of bases) in the antisense sequence (read in the 3' to 5' sense) is capable of pairing with the corresponding base (G with C, A with U) in the mRNA sequence read in the 5' to 3' sense. It is believed that this inhibition takes place by formation of a complex between the two complementary strands of RNA, thus preventing the formation of protein. How this works is uncertain: the complex may interfere with further translation, or degrade the mRNA, or have more than one of these effects. This antisense RNA may be produced in the cell by transformation of the cell with an appropriate DNA construct designed to transcribe the non-template strand (as opposed to the template strand) of the relevant gene (or of a DNA sequence showing substantial homology therewith).

The use of antisense RNA to downregulate the expression of specific plant genes is well known. Reduction of gene expression has led to a change in the phenotype of a plant, either at the level of gross visible phenotypic difference (e.g., lack of anthocyanin production in flower petals of petunia leading to colorless instead of colored petals (see van der Krol et al., *Nature*, 333:866-869 (1988)), or at a more subtle biochemical level, for example, a change in the amount of polygalacturonase

and reduction in depolymerization of pectin during tomato fruit ripening (Smith et al., *Nature*, 334:724-726 (1988)). Another more recently described method of inhibiting gene expression in transgenic plants is the use of sense RNA transcribed from an exogenous template to downregulate the expression of specific plant genes (Jorgensen, Keystone Symposium "Improved Crop and Plant Products through Biotechnology", Abstract X1-022 (1994)). Thus, both antisense and sense RNA have been proven to be useful in achieving downregulation of gene expression in plants, which are encompassed by the present invention.

Production of Transgenic Plants

Techniques for transforming a wide variety of higher plant species using the recombinant expression cassettes hereinbefore described are well known and described in the technical and scientific literature. See, for example, Weising *et al.*, *Ann. Rev. Genet.* 22:421-477 (1988).

The hereinbefore described recombinant expression cassettes may be introduced into the genome of a desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation, PEG poration, particle bombardment and microinjection of plant cell protoplasts or embryogenic callus, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. In the alternative, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes* host vector. The virulence functions of the *Agrobacterium* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

Transformation techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski *et al.*, *EMBO J.* 3:2712-2722 (1984). Electroporation techniques are described in Fromm *et al.*, *Proc. Natl.*

Acad. Sci. USA 82:5824 (1985). Biolistic transformation techniques are described in Klein *et al.*, *Nature* 327:70-73 (1987).

Agrobacterium tumefaciens-mediated transformation techniques are well described in the scientific literature. See, for example Horsch *et al.*, *Science* 233:496-498 (1984), and Fraley *et al.*, *Proc. Natl. Acad. Sci. USA* 80:4803 (1983). Although *Agrobacterium* is useful primarily in dicots, certain monocots can be transformed by *Agrobacterium*. For instance, *Agrobacterium* transformation of rice is described by Hiei *et al.*, *Plant J.*, 6:271-282 (1994).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the zmet2a and/or zmet2b methyltransferase nucleotide sequence(s). Plant regeneration from cultured protoplasts is described in Evans *et al.*, *Protoplasts Isolation and Culture, Handbook of Plant Cell Culture*, pp. 124-176, MacMillian Publishing Company, New York, 1983; and Binding; *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee *et al.*, *Ann. Ref. of Plant Phys.* 38:467-486 (1987)

The methods of the present invention are particularly useful for incorporating the zmet2a and/or zmet2b methyltransferase nucleic acid(s) into transformed plants in ways and under circumstances which are not found naturally. In particular, the zmet2a and/or zmet2b methyltransferase(s) may be expressed at times or in quantities which are not characteristic of natural plants.

One of ordinary skill in the art will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

The hereinbefore described expression cassettes can be inserted into a plant in order to reduce or alter the amount of DNA methylation in a plant. Preferably, such an expression cassette contains the zmet2a and/or zmet2b methyltransferase gene(s) inserted into the cassette in the antisense direction as described earlier. A reduction or alteration in the amount of DNA methylation in a plant can be used to stabilize transgene expression in a transgenic plant.

One of the difficulties with the production of transgenic plants is that many transgenes are silenced or are not stable through successive generations. In many cases, transgene silencing is associated with increased DNA methylation. The hereinbefore described expression cassettes of the present invention containing the zmet2a and/or zmet2b methyltransferase gene(s) in the antisense direction can be inserted into a plant either before, concurrently with or after the insertion of another expression cassette containing a transgene which is to be expressed in the plant, such as, but not limited to, a resistance or drought tolerance gene, etc. The antisense RNA produced by the hereinbefore described expression cassette can then form a complex with the endogenous mRNA from the zmet2a and/zmet2b methyltransferase gene(s) within the plant. This complex should reduce or alter the amount of DNA methylation occurring *in vivo* in the plant. This reduction in DNA methylation should prevent the silencing of the desired transgene in the plant.

In a similiar manner, the expression cassettes described herein can be used to modify or alter the yield or biochemical qualities of a plant. As discussed earlier, certain genes in plants and animals are expressed differentially when transmitted thorough a male versus female parent. This phenomenon is known as imprinting. Imprinting is an epigenetic system correlated with DNA methylation. A reduction or alteration of DNA methylation in a plant by transforming a plant with an expression cassette containing the zmet2a and/or zmet2b methyltransferase gene(s) in the antisense direction may affect the yield and biochemical qualities of a plant.

The hereinbefore described expression cassettes can also be used to silence the expression of a particular targeted gene in plants *in vivo*. More specifically, the

expression cassettes of the present invention containing a tissue-specific promoter and the zmet2a and/or zmet2b methyltransferase gene(s) in the sense direction can be inserted into a plant. The tissue-specific promoter will direct expression of the zmet2a and/or zmet2b methyltransferase gene(s) in a area containing the desired targeted gene. Translation of the zmet2a and/or zmet2b methyltransferase gene(s) in the specific area will result in an increase in methylation in the area of the targeted gene. This increase in methylation can silence the targeted gene.

Transgenic plants containing the expression cassettes described herein and which exhibit a reduction in DNA methylation can be identified by using methylation sensitive restriction enzymes or High Performance Liquid Chromatography. Techniques for using methylation sensitive restriction enzymes and High Performance Liquid Chromatography are well known in the art. Transgenic plants containing the expression cassettes described herein and which exhibit an increase in DNA methylation can be identified by using a Northern Blot analysis which is well known in the art.

Additionally, the hereinbefore described expression cassettes can be used in gene therapy for human diseases which are caused by the amplification of trinucleotide repeats.

The following Examples are offered by way of illustration, not limitation.

EXAMPLES

EXAMPLE 1 -Cloning and Sequencing of Zmet2a

a. Cloning and Sequencing

A partial cDNA clone (CGET064) from an immature tassel cDNA library was obtained from Pioneer Hi-Bred International (Des Moines, Iowa). This clone was identified in an expressed tag sequence (hereinafter "EST") database using known DNA methyltransferase sequences for comparison. This original cDNA clone contained sequences from bp 151 to bp 2569 shown in FIG. 1A and 1B. The sequence of this clone, which represents the 3' end of the transcript was used to design forward and reverse primers for 5' and 3' Rapid Amplification of cDNA Ends

(hereinafter "RACE"). RACE was conducted using the Marathon cDNA Amplification Kit (available from Clontech) on cDNA prepared from Mo17 10 day old seedling mRNA. Mo17 is publically available from the National Seed Storage Lab (Fort Collins, Colorado). RACE products were isolated and ends sequenced using Marathon primers and gene specific primers. The remaining sequence was obtained from PCR products by primer walking. The primers used were AP2, 1F, 1R, 2R, 3R, 4F, 5F, 8R, 8F, 9R, 9F, 14F, 17F, and RaceRT (see FIG. 3). Two sequencing passes were made on the Mo17 cDNA ends and four sequencing passes were made on the intervening regions, three from Mo17 cDNA and one from B73. B73 is publically available from the National Seed Storage Lab (Fort Collins, Colorado). A consensus sequence for the coding region was generated and is shown in FIG. 1A and 1B.

Genomic sequence spanning primers 1F and 1R were obtained from Pioneer Hi-Bred International. To obtain the remaining genomic sequence of zmet2a, the CGET064 clone was used to probe a Mo17 genomic library (Stratagene). Lambda clones 4a, 4c, 4d1 and 4d2 were determined to be positive clones containing a sequence identical to CGET064. Lamda clone 4a did not contain the full length gene, therefore, sequence data was obtained from clone 4c. No analysis of clones 4d1 or 4d2 was conducted. Clone 4c was subcloned into pGEM7zf(+) (Promega) using double digests involving *Hind*III, *Xho*I, *Eco*RI, and *Bam*HI. Genomic sequence was obtained from a combination of subclones pHX8 (bp 7311-8878), pHX9 (bp 9173-10135), and pB11(bp 5269-8447) and by primer walking using primers T7, Sp6, M13F, M13R, Seq2FN, Seq2RN, S3F, S3R, 7F, 8eR, 9F, 9R, 11iR, 11iF, 12iR, 12iF, 13iR, 13iF, 14F, 14R, 15R, 15F, 16R, 16F, 17R, 17F, 18R, 18F, and RaceRT (see FIG. 3). Borders of the *Mu* insertion of zmet2a::MU1 were sequenced from PCR products using primer 5F and a *Mu* primer (see FIG. 3). Map locations of the zmet2a primers are shown in FIG. 5.

PCR products were sequenced using Big Dye terminator cycle sequencing on an ABI sequencer (Perkin-Elmer Applied Biosystems) at the University of Wisconsin Biotechnology Center Sequencing Facility (Madison, WI). Sequence data was processed using computional tools available through the World Wide Web (hereinafter, "WWW"), summarized in FIG. 6.

b. Mutant Analysis

A mutant allele called (zmet2a::Mu1) was obtained from Pioneer Hi-Bred International's TUSC system. This mutant allele contains a *Mutator* transposable element insertion and was identified in a *Mutator* population using a *Mu* specific primer and a zmet2a gene specific primer. Since the *Mutator* population is quite variable, heterozygous zmet2a::mu1 F_2 seed was advanced by selfing at the University of Wisconsin West Madison Agronomy Farm (Madison, Wisconsin), the University of Wisconsin Walnut Street greenhouses (Madison, Wisconsin), and at the University of Wisconsin winter nursery in Puerto Rico to produce the F_4 derived F_5 segregating family primarily used in this example.

DNA from 15 plants of the F_4 derived F_5 segregating family was used for HPLC analysis. A subset of these plants was used for Southern analysis. The 5th to 7th immature leaf tips were collected and immediately frozen in dry ice. Tissue was ground in liquid nitrogen and DNA was extracted using a modified CTAB method of Saghai-Maroof et al. (*Proc. Natl. Acad. Sci. USA* 81:8014-8018 (1984)). Tissue was incubated in CTAB (Sigma) extraction buffer for 2 hours at 65 °C, extracted with chloroform/isoamyl alcohol, treated with 0.5 mg RNase A (Sigma) for 30 minutes at 37 °C, extracted again with choroform/isoamyl alcohol, precipitated with isopropanol, washed with 10mM ammonium acetate/76% ethanol, and resuspended in TE.

Plants were genotyped by Southern analysis. DNA (10µg) was digested with *Bam*HI and *Eco*RI which cut on each side of the *Mu* insertion. The digested DNA was electrophoresed through a 0.8% agarose 0.5X TBE gel. DNA was transferred to Immobillon nylon membrane (Millipore) with 5X SSC. Blots were UV cross-linked for 25 seconds and dried at 80 °C for 1.5 hours. Pre-hybridization was carried out in 5X SSC, 50mM Tris pH 8.0, 0.2% SDS, 10 mM EDTA, 2.5X Denhardts solution, and 0.1 mg/ml single stranded sheared herring DNA overnight (8-16 hours) at 65 °C. Hybridization conditions were similar to pre-hybridization except for the addition of 5% dextran sulfate to the hybridization solution. Probes (25-50 ng) (clone CGET064 for genotyping) were radioactively labeled using a random priming reaction

containing 50 µCi of P-32 labeled dCTP. Following overnight hybridization at 65 °C, blots were washed 2X (0.15X SSC, 0.1% SDS) for 30-45 minutes at 65 °C. Hybridized blots were then exposed to Kodak Biomax film.

Southern analysis with methylation sensitive restriction enzymes was conducted in a similar manner except that 5 µg of DNA was digested. Enzymes included in the study were: *Apa*I, *Ava*II, *Bam*HI, *Bgl*II, *Bst*NI, *Cla*I, *Eco*O109, *Eco*RI, *Eco*RII, *Hae*III, *Hin*fI, *Hha*I, *Hpa*II, *Msp*I, *Pst*I, *Pvu*II, *Sac*I, *Sau*3a, *Scr*F1, *Sma*I, *Xho*I. Probes for repetitive sequence regions of the maize genome including a 9 kb clone for the maize 26s-5.8s-17s repeat (reviewed in McMullen et al., *Molecular Analysis of the Nucleolus Organizer Region in Maize. In. Chromosome Engineering in Plants: Genetics, Breeding, and Evoluation*. Gupta PK, Tsuchiya T. (eds). pp. 561-576 (1991)), the 5s ribosomal subunit clone (Mascia et al., *Gene*, 15:7-20 (1981)), and centromere probe pSau3a9 (Jiang et al., *Proc. Natl. Acad. Sci. USA* 93:14210-14213 (1996)) were used to analyze changes in methylation due to zmet2a::Mu1.

HPLC was conducted according to a modified protocol of Gehrke et al., (*J Chromat.* 301:199-219 (1984)). Duplicate preparations for each of fifteen plants were analyzed. Twenty-five micrograms of DNA was diluted with water to a volume of 50 µl, denatured at 96 °C for 5 minutes and immediately placed on ice. One hundred microliters of 30mM ammonium acetate (pH 5.3), 5 µl of 20mM Zinc Sulfate and 10 µl Nuclease P1 (1mg/ml in 30mM ammonium acetate (pH 5.3) was added and incubated at 37 °C for 2 hours. This reaction cleaves 5' mononucleotides from single stranded DNA. The pH was adjusted with 20 µl of Tris (pH 8.5) and approximately 15 units of Calf Intestinal Alkaline Phosphatase was added and incubated at 37 °C for an additional 2 hours which converts the nucleotides to nucleosides. Samples were frozen at -20°C until HPLC analysis.

HPLC analysis was conducted at the University of Wisconsin Biotechnology Center. A volume of 50 µl was injected into a Brownlee Lab Spheri-5 RP-8 column. Nucleosides were separated with a flow rate of 0.75 ml/min using a gradient program consisting of 30 minutes in buffer A (0.05M Potassium Phosphate pH 4.0, 2.5% methanol), 19 minutes in buffer B (0.05M Potassium Phosphate pH 4.0, 20%

methanol). The column was flushed with 70% methanol for 13 minutes and then re-equilibrated with buffer A for 23 minutes before the injection of the next sample. All samples were analyzed on a Beckman System Gold chromatograph and nucleosides detected at A260nm and A280nm. Nucleoside and nucleotide standards (Sigma) were used to determine nucleoside peak positions and to create a standard curve to determine nucleoside concentration. The ratio of 5-methylcytosine to total cytosine was calculated and statistical analysis conducted using SAS.

To test remethylation as an indication of *de novo* methylase activity, an F_1 hybrid of an F_4 line homozygous for zmet2a::Mu1 and the inbred line Mo17 was backcrossed to the nonmutant Mo17 parent to generate plants homozygous wild-type and plants heterozygous for zmet2a::Mu1. Seedlings of the F_1, the BC_1 progeny, the Mo17 parent and a sib of the F_4 zmet2a::Mu1 parent were grown in the greenhouse and DNA extraction and Southern analysis conducted as previously described. DNA was digested with *Msp*I and *Pst*I and probed with the aforementioned repetitive clones.

c. Expression Analysis

The expression of zmet2a was determined by hybridizing the zmet2a cDNA probe to a Southern blot of cDNA's prepared from different tissues and tissues at different stages of development. Tissues included in this study are embryos 24 days after pollination, 10 day seedlings, immature ear, immature tassel, immature leaf from mutant and nonmutant plants, and roots. Total RNA was extracted using Trizol (Gibco/BRL) according to the manufacture's protocol. The PolyAttract System (Promega) was used to isolate mRNA's from all tissues except 10 day seedlings which was isolated using oligo dT cellulose columns (Pharmacia). cDNA was synthesized from the isolated RNA's using Marathon cDNA Amplification Kit (Clontech).

d. Results

zmet2a shares sequence similarity with other DNA methyltransferases

zmet2a is a member of a small gene family. Three cohybridizing bands are observed on a Southern blot of B73 DNA digested with *Hind*III and probed with

clone CGET064 which does not contain a *Hind*III restriction site (see FIG. 7). zmet2a, which maps to the long arm of chromosome 10, is coded on 20 exons with 19 intervening introns (FIG. 5). The inferred protein using the first predicted translation start site located within a consensus Kozak sequence (Kozak, *J. Cell. Biol.*, 115:887-903 (1991)) is composed of at least 912 amino acids with a predicted mass of 101 Kd (Kilodaltons). A protein of this size with an affinity for CpNpG sequences was isolated in *Pisum sativum* by Pradhan and Adams (*Plant J.*, 471-481 (1995)).

Comparisons with *Arabidopsis* chromomethylase, *CMT1*

Sequence of zmet2a (FIG. 1A and 1B) reveals that it lacks the large N-terminal domain found in the maintenance enzymes but does possess the six highly conserved motifs of the C-terminal catalytic domain. Database searches using BLAST (http://www.ncbi.nlm.nih.gov/gov/BLAST/) show that zmet2a has highest sequence homology to the *Arabidopsis* chromomethylase, *CMT1* (see Henikoff and Comai, *Genetics*, 148:307-318 (1998)) with 44% identity, 57% conservation. The N-terminal region is larger in zmet2a; however, there is an additional downstream predicted start site, also within a consensus Kozak sequence, that codes for an enzyme of 809 amino acids which is more similar in size to the most closely related *CMT1* which is composed of 791 amino acids.

Alignments of zmet2a with *CMT1* and the catalytic domains of *Arabidopsis* *MET1* and maize *zmet1* maintenance enzymes show conservation in the important functional motifs I, IV, VI, VIII, IX and X providing evidence that it is indeed a DNA methyltransferase (FIG. 8). zmet2a and *CMT1* are 87% conserved across the defined six conserved domains, as shown in the underlining in FIG. 8. Zmet2a and CMT1 also have 60% conservation in the variable region sequence between the defined underlined motifs VIII and IX in FIG. 8, which contains a region known as the target recognition domain in the bacterial methyltransferases. The bacterial methylase M.*Hha*I has been crystalized and functions deduced for the conserved amino acids (Cheng et al., *Cell*, 74:299-307 (1993)). The zmet2a amino acids involved in catalysis were predicted by comparison to M.HhaI. The amino acids interacting with SAM and with cytosine are summarized in FIG. 9.

zmet2a mutant plants have reduced methylation at CpNpG sites

A reverse genetics approach was used to ascertain the function of zmet2a. A F_2 family segregating for a *Mutator (Mu)* insertion in the exon encoding motif IX was identified using a PCR primer for *Mu* and a gene-specific primer for zmet2a. This allele is called zmet2a::Mu1. The insertion of *Mu* into exon 19 results in a transcript that would code for a protein truncated at the point of the *Mu* insertion in motif IX due to the introduction of a stop codon. The resulting protein is expected to be dysfunctional since it lacks Motif X which is required for S-Adenosyl methionine (hereinafter "SAM") binding (Cheng et al. *Cell*, 74:299-307 (1993)).

Reduced methylation observed by restriction enzyme analysis

To reduce the genetic background variation associated with the heterogeneous origin of the *Mutator* population, restriction enzyme analysis was conducted on a F_4 derived F_5 family segregating for zmet2a::Mu1. Restriction enzyme isoschizomers *Hpa*II/*Msp*I in addition to other methylation sensitive enzymes were used to determine methylation pattern differences among the three genotypic classes. *Hpa*II and *Msp*I both recognize the sequence CCGG but differ in their sensitivity to methylation. *Hpa*II digestion is inhibited unless both cytosines are unmethylated whereas *Msp*I can digest $C^{me}CGG$ sequences but not $^{me}CCGG$ sites. The methylation status at CpG sites can be accessed by digesting with *Hpa*II and similarly *Msp*I digestion is used to determine the state of methylation at CpCpG sites specifically and may provide a general indication of methylation changes occurring at CpNpG sites.

Results indicate significant reductions in cytosine methylation at ^{me}CCG sites as indicated by a more complete digestion by *Msp*I in plants homozygous for zmet2a::Mu1 (FIG. 10 A-C). Plants heterozygous for zmet2a::Mu1 were intermediate in their digestion pattern. Although the frequency of methylated cytosines is much higher at CpG sequences, no changes in methylation were observed among the genotypic classes when digested with *Hpa*II (FIG. 10 A-C).

Isoschizomers, *Bst*NI and *Eco*RII recognize the sequence CC(A/T)GG. *Bst*NI is not sensitive to cytosine methylation and *Eco*RII is inhibited at $C^{me}C(A/T)GG$ sites. Nearly all of these sites are methylated in repetitive sequences as a low level of

*Eco*RII digestion is observed only in zmet2a::Mu1 plants (See FIG. 11), whereas digests with *Bst*NI are completely digested to lower molecular weight fragments for all genotypes. These methylated sites may not be subject to zmet2a activity but may instead be methylated by another member of the zmet2a gene family or by *zmet1* or possibly *de novo* methylated after each cell cycle by *zmet3*. Other restriction enzymes were used to clarify the apparent sequence specificity of methylation reduction at CpNpG sites. As with the isoschizomers, no digestion differences are observed with CpG sensitive enzymes *Hha*I [G^{me}CGC] and *Cla*I [AT^{me}CGAT]. More complete digestion is observed in plants homozygous for zmet2a::Mu1 with enzymes sensitive to methylation at CpNpG sites. FIG. 12 shows digestion patterns for enzymes sensitive to methylation at CpNpG sites: *Eco*RII, *Bgl*II, *Pst*I, *Bam*HI, and *Ava*II. In addition to *Eco*RII as previously mentioned, reduced methylation in one or more of the repetitive regions was observed with *Bgl*II [AGAT^{me}CT], *Pst*I [^{me}CTGCAG], *Bam*HI [GGAT^{me}CC], and *Ava*II [GG(A,T)^{me}C^{me}C]. It should be noted that *Ava*II may include some CpG overlapping sites. Subtle differences in digestion patterns of one or more of the repetitive sequences were also observed with *Sau3a*I [GAT^{me}C], *Apa*I [GGG^{me}CC^{me}C], and XhoI [^{me}CT^{me}CGAG]. With these enzymes it is not possible to unambiguously determine whether the source of the difference is CpG or CpNpG methylation. Differences were also observed with *Scr*FI [C^{me}CNGG] which duplicates the targeted sequences and methylation sensitivities of *Eco*RII, *Msp*I and *Hpa*II. Although in many cases the observed reduction in CpNpG or CpN methylation is minimal, any cases of reduced methylation that could be unambiguously attributed to CpG sites have not been observed.

Reduced methylation observed by HPLC

To further assess the extent of methylation reduction caused by the zmet2a::Mu1 allele, HPLC was used to determine the proportion of methylated cytosines in the same F_5 plants used for restriction enzyme analysis. An 11.6% decrease in 5-methylcytosine was observed in plants homozygous for zmet2a::Mu1 relative to siblings homozygous for wild-type zmet2a (FIG. 12). Heterozygotes were intermediate in 5-methylcytosine content. Differences between the genotypic classes are statistically significant at $\alpha < 0.0001$. Since most methylation is found at CpG sites (Gruenbaum et al., *Nature*, 292:860-862 (1981)), a 12% decrease in the total 5-

methylcytosine content likely accounts for a substantial reduction in methylation at CpNpG sites if the reductions are confined to these sequences.

Several generations of inbreeding does not reduce methylation levels beyond that which is observed in the F_2 homozygous mutant (FIG. 13). In addition, it was also observed that plants restored to a normal zmet2a genotype from zmet2a::Mu1 heterozygotes appeared to have near normal levels of methylation.

Methylation is restored after segregation away from zmet2a::Mu1

To test remethylation, a nonmutant line, Mo17, was crossed to a homozygous mutant line, the resulting F_1 was then backcrossed to the nonmutant Mo17 parent line. Restriction enzyme analysis of backcross progeny show all individuals without the *Mu* insertion have remethylated to levels similar to the backcross parent (see FIG. 14). The increased levels of methylation observed in normal BC_1 progeny appear to be higher than that expected from the segregation of normal Mo17 derived chromosome segments and low methylation mutant segments, which would result in a pattern intermediate between the F_1 and the nonmutant parent. These results indicate either that zmet2a has *in vivo de novo* activity and is responsible for establishing CpNpG methylation patterns, or that a separate *de novo* methyltransferase functions only early in development and that zmet2a is responsible for maintaining these patterns. These results on remethylation are in contrast to those of the reduced methlyation patterns of *Arabidopsis* mutants. Backcross progeny, lacking an antisense *MET1* transgene or the *ddm1* mutation, derived from mutant plants outcrossed to normal plants showed very slow remethylation and required several generations to restore methylation to normal levels (Ronemus et al., *Science*, 273:654-657 (1996), Vongs et al., *Science*, 260:1926-1928 (1993), Kakutani et al., *Genetics*, 151:831-838 (1999)). Similar results were observed in selfed progeny from hemizygous antisense *Met1* plants that did not inherit the transgene (Finnegan et al., *Proc. Natl. Acad. Sci. USA* 93:8449-8454 (1996)) however a centromeric region and some single copy sites did remethylate in the first generation (Finnegan et al., *Annu. Rev. Plant Physiol. Plant Mol. Bio.*, 49:223-247 (1998)).

Other DNA methyltransferases that lack the large N-terminal domain have been presumed to be *de novo* enzymes, however, evidence remains insufficient. *In vitro* expression of *Dnmt*3a and *Dnmt*3b (Okano et al., *Nature Genetics*, 19:219-220 (1998)) did not show a specific preference for hemimethylated DNA or nonmethylated DNA and in vivo expression in *Drosophila* (Lyko et al., *Nature Genet.*, 23:363-366 (1999)) further confirm *de novo* activity, whereas *Dnmt*2 (Okano et al., *Nucleic Acids Res.*,26:2536-2540 (1998)) was shown not to effect *de novo* or maintenance methylation in mice. *Masc1*, in *ascobolus*, is purported to have *de novo* activity through its effect on methylation induced premeiotically (MIP) (Malagnac et al., *Cell*, 91:281-290 (1997)). Another Ascobolus methyltransferase *Masc2* was found to be dispensible for maintenance and de novo methylation *in vivo* (Malagnac et al, *Mol. Micro.* 3:331-338 (1999)).

A chromodomain is present in zmet2a

A distinguishing feature of zmet2a, like *CMT1*, is the presence of the chromodomain. Chromodomains have been demonstrated to target proteins to heterochromatic regions and may also be a site of protein-protein interactions (reviewed by Cavalli and Paro, *Curr. Op. Cell Biol.*, 10:354-360 (1998)). The presence of the chromodomain in zmet2a and *CMT1* potentially suggests targeting of the methyltransferase to chromatin complexes or a role of the methyltransferase in chromatin formation and stability. Furthermore, the observation that zmet2a affects CpXpG methylation may also implicate protein targeting through the chromodomain and targeting of methylation patterns. Stable transcriptionally active or silent states may be determined by the formation of chromatin complexes. The mechanisms involved in the formation of silencing complexes remain unknown. However, there is evidence of the involvement of methylation in transcriptionally silenced states which involve methylation binding proteins, transcriptional repressor complexes, and histone deacetylases (Nan et al., *Nature*, 393:386-389 (1998), Wade et al, *Nature Gen.*, 23:62-66 (1999), Ng et al., *Nature. Gen.* 23:58-61 (1999)).

zmet2a is expressed throughout plant development. Expression is higher in the rapidly dividing tissues of seedling, immature ear and embryos (FIG. 15) consistent with the role of methyltransferases in methylating newly synthesized DNA.

Low expression of zmet2a in terminal tissue (leaves) could serve a protective function against invading DNA if this enzyme does have a *de novo* function.

Example 2 – Cloning and Sequencing of the maize retrotransposon SPRITE-1

This example describes the cloning and sequencing of a maize retrotransposon that is inserted into an intron of zmet2a and is referred to herein as "SPRITE-1".

a. Introduction

Within the genomes of most organisms are DNA elements that can be considered parasitic. These elements confer no phenotype of their own and function only for their propagation and insertion elsewhere in the genome. There are two major classes of these elements based on the mechanisms of propagation. One class propagates using DNA-mediated mechanisms where the element does not code for any polymerase and entirely depends on the replication machinery of the host. This class includes the *Ac*, *Spm*, and *Mu* transposable element systems. The other major class is known as retrotransposons, retrotransposable elements or retroelements (reviewed in Grandbastien, *Trends in Genetics* 8:103-108 (1992); Eickbush, *Origin and Evolutionary Relationships of Retroelements. In The Evolutionary Biology of Viruses* (Morse, S.S.. ed).) (1994); Wessler et al., *Current Biology*, 5:814-821 (1995); Bennetzen, *Genome*, 37:565-576 (1996)). These elements are not able to excise from one site and insert into another, as the previously mentioned class is capable, but replicate by an RNA-mediated process. The retroelements code for a reverse transcriptase which is a DNA polymerase that uses RNA as a template.

There are several types of retroelements. The main types are retroviruses, long-terminal-repeat (hereinafter "LTR") retroelements, and non-LTR retroelements. Retroviruses are infectious and have not been found in plants, although one plant LTR-retroelement, SIRE-1 from soybean has coding sequences similar to that of a retoviral envelope protein (Laten et al., *Proc. Natl. Acad. Sci.*, 95:6897-6902 (1998)). The non-LTR class is mainly composed of long interspersed nuclear elements (hereinafter "LINEs") and short interspersed nuclear elements (hereinafter "SINEs"). These elements have been found in plants. Less is known about this class than the others. They do differ from LTR-retroelements in that they contain a poly-A tail at

their 3' end. The LTR-retroelement class has been more extensively described in plants than the other classes of retroelements. The LTR-retroelements are usually categorized as one of two groups based on the similarity with the first elements described in yeast and *Drosophila*. One group shares similarity with the Ty3 elements from yeast and the *gypsy* element of *Drosophila* (Marlor et al., *Mol. Cell. Biol.*, 22:829-846 (1986); Clark et al., *J. Biol. Chem.*, 263:1413-23 (1988)). The other group has similarity with the Ty1 elements of yeast and the *copia* element of *Drosophila*. The element identified in this study is of the Ty1/*copia* class (Clare and Farabaugh, *Proc. Natl. Acad. Sci. USA*, 82:2829-2833 (1985); Mount and Rubin, *Mol. Cell. Biol* 5:1630-1638 (1985)).

The general structure of a LTR-retroelement is depicted in FIG. 16A. These elements are similar in their structure and replication to retroviruses (reviewed in Witcomb and Hughes, *Ann. Rev. Cell Biol.*, 8:275-306 (1992), Eickbush, *Origin and Evolutionary Relationships of Retroelements.* In *The Evolutionary Biology of Viruses* (Morse, S.S., ed.). New York: Raven Press, pp 121-157 (1994), Bennetzen, *Trends in Microbiology*, 9:347-353 (1996)). These elements have direct repeats at the termini as opposed to the DNA based elements that have inverted terminal repeats. Downstream from the 5' LTR is a primer binding site for a host tRNA that primes the first DNA strand synthesis using reverse transcriptase. One or more open reading frames that code for *gag*, a protease, an integrase, a reverse transcriptase, and RNaseH are located downstream from the primer binding site. After the coding region is a polypurine tract followed by the 3' LTR. Ty3/*gypsy* and Ty1/*copia* elements differ in the postion of the integrase coding region. Ty3/gypsy element have the integrase domain at the end of the coding region whereas Ty1/*copia* element have it positioned between the proteinase and reverse transcriptase regions. The *gag* gene encodes proteins for the nucleocapsid and the highly conserved cysteine-histidine nucleic acid binding domain ($CX_2CX_4HX_4C$). The protease processes the polyprotein into its individual components. The integrase functions to insert a newly replicated element into the host DNA. The reverse transcriptase synthesizes the first DNA strand from the transcribed RNA of the element. The RNase degrades the RNA following first strand synthesis. Retroelements rely on the RNA polymerase of the host for

transcription and the host DNA polymerase for second strand DNA synythesis to complete replication.

Using PCR based methods, retroelements were found within nearly every species of the plant kingdom studied (Flavell et al., *Nuc. Acids Res.* 20:3639-3644 (1992); Voytas et al., *Proc. Natl. Acad. Sci. USA* 89:7124-7128 (1992)). Despite the ubiquitous nature of retroelements, there is great heterogeneity among the element within and among species (Flavell et al., *Nuc. Acids Res.* 20:3639-3644 (1992), Wang et al., *Plant Mol. Biol.*, 33:1051-1058 (1997), Pearce et al., *Mol. Gen. Genet.*, 250:305-315 (1996)).

Retroelements are found to be distributed over the entire lengths of chromosomes in *Avena sativa* (Katsiotis et al., *Genome*, 39:410-417 (1996)) but have also been found to be less abundant in heterochromatin, nucleolar organizer regions, centromeres and telomeres (Pearce et al., *Mol. Gen. Genet.*, 250:305-315 (1996); Moore et al., *Genomics*, 10:469-476 (1991); Aledo et al., *Theor. Appl. Genet.*, 90:1094-1100 (1995); Brandeis et al., *Plant Mol. Biol.*, 33:11-21 (1997)). Retroelement-like sequence were found in centromeric regions of grass chromosomes (Miller et al., *Genetics*, 150:1615-1623 (1998)). Many retroelements were discovered by their associations with plant genes (Johns et al., *EMBO J.*, 4:1093-1102 (1985); Grandbastien et al., *Nature*, 337:376-380 (1989); Camirand et al., *Mol. Gen. Genet.*, 224:33-39 (1990)); White et al., *Proc. Natl. Acad. Sci. USA*, 91:11792-11796 (1994)); Hu et al., *Mol. Gen. Genet.*, 248:471-480 (1995); Bi and Laten, *Plant Mol. Biol.*, 30:1315-1319 (1996), Royo et al., *Mol. Gen. Genet.*, 250:180-188 (1996); Kumekawa et al., *Mol. Gen. Genet.*, 260:593-602 (1999)). Many more retroelements or retroelement fragments have been identified using PCR with degenerate primers (Voytas et al., *Proc. Natl. Acad. Sci. USA*, 89:7124-7128 (1992)); Flavell et al., *Nuc. Acids Res.*, 20:3639-3644 (1992); Flavell et al., *Mol. Gen. Genet.*, 231-233 (1992), Pearce et al., *Mol. Gen. Genet.*, 250:305-315 (1996); Katsiotis et al., *Genome*, 39:410-417 (1996); Wang et al., *Plant Mol. Biol.*, 33:1051-1058 (1997)). Others have been identified through studies for other purposes (Bhattacharyya et al., *Plant Mol. Biol.*, 34:255-264 (1997); Vicient and Martinez-Izquierdo, *Gene*, 184:257-261 (1997);

Manninen and Schulman, *Plant Mol. Biol.*, 22:829-846 (1993)) or by genome sequencing projects.

The Ty3/*gypsy* and the Ty1/*copia* elements can be found in large numbers and may contribute up to 50% of the nuclear DNA of the maize genome (SanMiguel et al., *Science*, 274:765-768 (1996)). A 280 Kb region of the maize genome containing the Adh1-F and u22 genes was composed of retroelements, from 10 different families, inserted within each other. The copy number of Ty1/*copia* elements varies considerably. For example, the Ta1 elements of *Arabidopsis* (Voytas et al., *Genetics*, 126:713-721 (1990)) and the Tst1 element of *Solanum tuberosum* (Camirand et al., *Mol. Gen. Genet.*, 224:33-39 (1990)) have one to only a few copies whereas the maize element PREM-2 (Bennetzen, *Trends in Microbiology*, 9:347-353 (1996)) and the BARE-1 element of *Hordeum vulgare* (Manninen and Schulman, *Plant Mol. Biol.*, 22:829-846 (1993)) may be present at 30.000 or more copies.

The differences in copy number infer differences in expression of retroelements. Retroelements are not expressed at high levels as only a few examples of activity have been observed. The Bs1 and Zeon-1 elements of maize (Johns et al., *EMBO J.*, 4:1093-1102 (1985); Hu et al., *Mol. Gen. Genet.*, 248:471-480 (1995)); the Tos elements of rice (Hirochika et al., *Proc. Natl. Acad. Sci. USA* 93:7783-7788 (1996)) the Tnt1 and Tto1 elements of tobacco (Grandbastien et al., *Nature*, 337:376-380 (1989); Hirochika, *EMBO J.*, 12:2521-2528 (1993)) and the Tnp2 element of *Nicotiana plumbaginifolia* have shown evidence of activity. Retroelement expression is higher in plant tissues under stressful conditions. The Tto1, Tto2 of tobacco and Tos17 element of rice were shown to be activated in tissue culture (Hirochika, *EMBO J.*, 12:2521-2528 1993, Hirochika et al., *Proc. Natl. Acad. Sci.*, USA (1996)). The promotors of the BARE-1 element of barley and the Tnt-1 element of tobacco drove expression of reporter genes in protoplasts (Suoniemi et al., *Plant Mol. Biol.*, 31:295-306 (1996); Pouteau et al., *EMBO J.*, 10:1911-1918 (1991)).

Biotic stresses such as viral, fungal and bacterial infection and abiotic stress such as wounding have also been shown to initiate the expression of Tnt1 and Tto1 retroelements (Pouteau et al., *Plant J.*, 5:535-542 (1994); Moreau-Mhiri et al., *Plant*

J., 9:409-419 (1996); Vernhettes et al., *Plant Mol. Biol.*, 35:673-679 (1997); Mhiri et al., *Plant Mol. Biol.*, 33:257-266 (1997); Grandbastien et al, *Genetica*, 100:241-252 (1997); Takeda et al., *Plant Mol. Biol.*, 36:365-376 (1998)). The Bs1 element of maize may have been mobilized prior to insertion in the Adh1 gene by infection with the barley stripe mosaic virus (Johns et al., *EMBO J.*, 1093-1102 (1985)). Only the expression of BARE-1 has been observed in normal unstressed barley leaves (Suoniemi et al., *Plant Mol. Biol.*, 31:295-306 (1997)).

Under normal conditions, retroelements are transcriptionally inactive and are thus transpositionally inactive. Mechanisms within the host must exist to regulate the activity of the retroelements to prevent potentially deleterious mutations that could occur if retroelement transposition was unchecked. Most retroelements are highly methylated (Bennetzen et al., *Genome*, 37:565-576 (1994)) and possibly in heterochromatic regions and may not be accessible to transcriptional machinery. Though silenced in most cases and active in stressful situations, it has been suggested that retroelement transposition may create mutations that may be of selective advantage and provide a means for adaptation (McClintock, *Science*, 226:792-801 (1984)).

b. Cloning and Sequencing of SPRITE-1.

A zmet2a genomic clone was isolated from a lambda library (Stratagene) constructed from Mo17 genomic DNA. The sequence was obtained from subclones or from PCR products by primer walking. Fragments were sequenced using Big Dye terminator cycle sequencing on an ABI sequencer (Perkin-Elmer Applied Biosystems) at the University of Wisconsin Biotechnology Center Sequencing Facility, Madison, Wisconsin.

Expression analysis was conducted on cDNA's prepared using Marathon cDNA Amplification Kit (Clontech) according to the manufacturer's protocols from mRNA isolated from a Mo17 10 day old seedling, Mo17 immature tassel, B73 immature ear, Black Mexican Sweet (BMS) callus, Mo17 embryo 24 days after pollination, W22 pollen, young roots, and immature leaf tissue from zmet2a normal and mutant plants. Total RNA was extracted using Trizol (Gibco/BRL) according to

manufacturer's protocol. Seedling mRNA was isolated using oligo dT cellulose columns (Pharmacia) all other mRNA isolated using the PolyAttract system (Promega).

c. DNA extraction and Southern analysis for genotyping and methylation analysis.

DNA was extracted from immature leaf blades as described in Saghai Maroof et al. (*Proc Natl. Acad. Sci. USA* 81:8014-8018 (1984)). The copy number of SPRITE-1 was determined by digesting DNA (10μg) with *Eco*RI which does not cut within the element. The digested DNA was electrophoresed through a 0.8% agarose 0.5X TBE gel. Gels were treated with 0.25N HCl for 15 minutes, denatured in 0.2N NaOH and 0.6 M NaCl for 30 minutes, then neutralized in 0.5 M Tris 1.5 M NaCl for 30 minutes. DNA was transferred to Immobilon nylon membrane (Millipore) with 5X SSC. Blots were dried at 80 °C for 1.5 hours. Pre-hybridization was carried out in 5X SSC, 50 mM Tris pH 8.0, 0.2% SDS, 10 mM EDTA, 2.5X Denhardt's solution, and 0.1 mg/ml single stranded sheared herring DNA overnight (8-16 hours) at 65 °C. Hybridization conditions were similar to pre-hybridization except for the addition of 5% dextran sulfate to the hybridization solution. The blot was probed with a PCR fragment (25-50 ng) amplified from the 5' end of the element. Probes were P-32 (50 μCi) labeled using random priming. Following overnight hybridization at 65 °C, blots were washed 2X (0.15X SSC, 0.1% SDS) for 30-45 minutes at 65 °C. Hybridized blots were then exposed to Kodak BioMax film. Southern analysis with methylation sensitive restriction enzymes was conducted on B73 and Mo17 using the same protocols as for genotyping except that 5 μg of DNA was digested. Enzymes included in the study were the differentially methylation sensitive isoschizomers *Hpa*II/*Msp*I and *Eco*RII/*Bst*NI as well as other methylation sensitive enzymes: *Hha*I, and *Pst*I. Blots were hybridized with probes representing different portions of the element.

d. HPLC analysis.

HPLC was conducted according to a modified protocol of Gehrke et al. (*J. Chromato.*, 301:199-219 (1984)). B73 x Mo17 recombinant inbred lines carrying a SPRITE-1 insertion were determined using PCR with the zmet2a primers 15F and 8R,

and the SPRITE-1 primer 18R. Preparations for each of four plants with and without SPRITE-1 were analyzed. Twenty-five micrograms of DNA was diluted with water to a volume of 50 µl, denatured at 96 °C for 5 minutes and immediately placed on ice. One hundred microliters of 30 mM ammonium acetate (pH 5.3), 5 µl of 20 mM Zinc Sulfate and 10 µl Nuclease P1 (1mg/ml in 30 mM ammonium acetate (pH 5.3) was added and incubated at 37 °C for 2 hours. This reaction cleaves 5' mononucleotides from single stranded DNA. The pH was adjusted with 20 µl of Tris (pH 8.5) and approximately 15 units of Calf Intestinal Alkaline Phosphatase was added and incubated at 37 C for an additional 2 hours which converts the nucleotides to nucleosides. Samples were frozen at -20 °C until HPLC analysis.

HPLC analysis was conducted at the University of Wisconsin Biotechnology Center, Madison, Wisconsin. A volume of 40 µl was injected into a Brownlee Lab Spheri-5 RP-8 column. Nucleosides were separated with a flow rate of 0.75 ml/min using a gradient program consisting of 30 minutes in buffer A (0.05M Potassium Phosphate pH 4.0, 2.5% methanol), 19 minutes in buffer B (0.05M Potassium Phosphate pH 4.0, 20% methanol). The column was flushed with 70% methanol for 13 minutes and then re-equilibrated with buffer A for 23 minutes before the injection of the next sample. All samples were analyzed on a Beckman System Gold chromatograph and nucleosides detected at A260 nm and A280 nm. Nucleoside and nucleotide standards (Sigma) were used to determine nucleoside peak positions and to create a standard curve to determine nucleoside concentration. The ratio of 5-methylcytosine to total cytosine was calculated and statistical analysis conducted using SAS.

e. Expression analysis.

The expression of SPRITE-1 was determined by hybridizing a SPRITE-1 probe to a Southern blot of cDNA's prepared from different tissues and tissues at different stages of development. Tissues included in this study are embryos 24 days after pollination, 10 day seedlings, immature ear, immature tassel, immature leaf from mutant and nonmutant plants, roots, BMS callus, and mature pollen. Total RNA was extracted using Trizol (Gibco/BRL) according to the manufacture's protocol. The PolyAttract System (Promega) was used to isolate mRNA's from all tissues except 10

day seedlings which was isolated using oligo dT cellulose columns (Pharmacia). cDNA was synthesized from the isolated RNA's using Marathon cDNA Amplification Kit (Clontech).

f. Results

SPRITE-1 is similar to retrotransposons of the Ty1/*copia* group.

In the process of sequencing the maize methyltransferase gene zmet2a, a retroelement inserted within an intron of this gene was discovered and names SPRITE-1. This element is positioned in opposite transcriptional orientation relative to zmet2a. The insertion spans 5220 bp and possesses all the components of a retroelement. Sequence data indicates that SPRITE-1 is a Long-Terminal-Repeat (hereinafter "LTR") retroelement belonging to the Ty1/*copia* class of retroelements. FIG. 16a depicts the general structural components of SPRITE-1. FIG. 16b shows the sequence of the terminal structural components. SPRITE-1 has a perfect 109 bp direct terminal repeats which includes a 3 bp inverted repeat that flanks the internal element sequence. These repeats have the TG...CA pattern found in most plant retroelements and are also shorter than LTR's of most retroelements. LTR's range in size from 115 bp to 4560 bp from information compiled by Bennetzen (*Trends in Microbiology*, 9:347-353 (1996)). A 5 bp host site duplication flanks the repeats externally. Downstream and adjoining the 5' LTR is a primer binding site (PBS) of 16 bp that has sequence complementary to the wheat germ cytoplasmic initiator methionine tRNA (Ghosh et al., *Nuc. Acids. Res.*, 10:3241-3247 (1982)). Upstream and adjoining the 3' LTR is a polypurine tract of 9 bp. Between the putative transcription start site to the predicted translation start site is a 550 bp untranslated region. SPRITE-1 contains a single open reading frame coding 1485 amino acids ending with the stop codon at the 5' end of the polypurine tract.

Database searches for similar coding sequences using BLAST (http://www.ncbi.nlm.nih.gov/gov/BLAST/) show that SPRITE-1 belongs to a different family of retroelements than any other previously described. The most closely related elements based on overall amino acid similarity include an *Arabidopsis* retroelement (AC006528), Retrofit from *Oryza longstaminata* (U72725), and Hopscotch from *Zea mays* (U12626) all having ~35% identity and ~50%

conservation in amino acid sequence with SPRITE-1. It also shares 29% identity and 45% conservation with the *copia* element from *Drosophila*. No elements were found to have nucleotide similarity with the LTR of SPRITE-1 further indicating that this is a member of a unique family of Ty1/*copia* type elements.

SPRITE-1 has the component retrovirus-like amino acid motifs that code for the proteins necessary for transposition. These motifs are the gag-related protein that contains a Cys-His box also known as the CCHC zinc-binding domain, a protease, an integrase, reverse transcriptase and RNase H. These motifs are ordered as they are in Ty1 and *copia*. FIG. 17 shows amino acid alignments of these conserved region from the similar retroelements previously mentioned. These motifs were similarly positioned relative to each other in these retroelements except the CCHC zinc binding domain which was more variant in position relative to the protease motif. This motif was aligned by hand whereas the alignments of the other motifs were constructed by CLUSTAL W and processed using BOXSHADE. Alignments indicate that SPRITE-1 does possess the component protein coding regions necessary for replication and transposition. The coding regions of many retroelements have shown mutations that create frameshifts or introduce stop codons thus preventing translation of functional proteins and preventing transposition. The coding region of SPRITE-1 is intact and therefore has the potential to transpose.

The number of copies of SPRITE-1 is relatively low but variable.

A survey of inbred lines developed from several different populations and other genetic stocks revealed differences in SPRITE-1 copy number. DNA was digested with *Eco*RI and southern blots hybridized with a probe representing the 5' untranslated region of SPRITE-1. This element does not have any *Eco*RI restriction sites. SPRITE-1 is found at a low copy number in most maize lines. Copy number varies from 3 as in B73 and Mo17 to 5 as in B14 and B79 (FIG. 18). The insertion of SPRITE-1 into zmet2a is only found in Mo17 and not in any other maize inbred line except A682, a line derived from Mo17 (FIG. 19). C.I. 187-2, a Mo17 parental line, does not contain SPRITE-1. This indicates that SPRITE-1 has been active recently, i.e. after the origin of the maize populations used for inbred development.

Expression of SPRITE-1

Expression was investigated by hybridizing a southern blot of cDNAs, sythesized from mRNA from different maize tissues, with a SPRITE-1 probe (FIG. 20). Expression of SRITE-1 was highest in leaf tissue. Expression was highest in leaf tissue from plants with a *MUTATOR* insertion in zmet2a and decreased CpNpG methylation. A low level of expression was observed in most tissues, but this may be due to transcription of other genes containing SPRITE-1 in a sense orientation.

SPRITE-1 does not effect zmet2a transcript processing.

During the sequencing of zmet2a cDNA, no fragments or subclones possessed SPRITE-1 sequence indicating that it is efficiently spliced from the transcript. Aberrant splicing has been observed in genes containing retroelements (Pouteau et al., *Mol. Gen. Genet.*, 228:233-239 (1991), Varagona et al., *Plant Cell*, 4:811-820 (1992), Marillonnet and Wessler, *Plant Cell*, 9:967-978 (1997), Kapitonov and Jurka, *J. Mol. Evol.*, 48:248-251 (1999)). Expression of three alleles of the *waxy* gene of maize was low due to retroelement insertions within introns (Varagona et al., *Plant Cell*, 4:811-820 (1992)). Varagona et al. (*Plant Cell*, 4:811-820 (1992)) found that although the element was spliced out of the *waxy* transcript, long-range splice site recognition was disrupted as exons upstream and downstream of the insertion site were found to be excluded in some transcripts. Further analysis of the *wxG* allele showed tissue specific differences in RNA processing with more correctly spliced transcripts in pollen than in the endosperm (Marillonnet and Wessler, *Plant Cell*, 9:967-978 (1997)).

Alternatively spliced transcripts were searched for by PCR amplification of fragments spanning several exons both upstream and downstream of the SPRITE-1 insertion site. Fragments were amplified from Mo17 seedling and immature embryo cDNA and compared to fragments amplified from B73 immature ear cDNA (FIG. 21). Amplification products were separated on an agarose gel and southern blotted. The Southern blot was hybridized to a near full length zmet2a cDNA. No differences were observed between the B73 and Mo17 products indicating that only correctly spliced fragments were detected. The blot was stripped and probed with retroelement sequences. No transcripts were amplified that contained any SPRITE-1 sequence. In

the tissues examined in this example, no aberrant transcripts were detected. Aberrant splicing products may be at such a low concentration that they are not detectable.

SPRITE-1 does not effect zmet2a expression and function.

Since SPRITE-1 is inserted into an intron of zmet2a, the effect of this insertion on zmet2a activity was investigated. HPLC data shows no methylation differences among the recombinant inbred lines with or without a SPRITE-1 insertion in zmet2a. Lines with a SPRITE-1 insertion had 18.21% ± 1.78 5-methylcytosine whereas lines without the insertion had 18.20% ± 0.24. It is probable that most transcripts are processed correctly since no changes in methylation are observed in plants with a SPRITE-1 insertion.

Regions of SPRITE-1 are hypermethylated

Portions of SPRITE-1 were examined to determine the status of cytosine methylation. Using methylation sensitive restriction enzymes, sites within 970 bp of the untranslated region (hereinafter "UTR") immediately downstream from the transcription start site was analyzed. FIG. 22 shows methylation sensitive restriction digestion patterns for Mo17 and B73. The isoschizomers *Hpa*II and *Msp*I recognize CCGG sequences and are differentially sensitive to methylation. SPRITE-1 has a single *Msp*I/*Hpa*II site. Using the SPRITE-1 sequence from Mo17, the zmet2a insertion of SPRITE-1 would generate fragments of 5853 bp and 4625 bp. Other SPRITE-1 insertions would generate fragments of variable lengths. Southern blots show only very large fragments >20 Kb for both *Hpa*II and *Msp*I. *Msp*I does show a smaller fragment size than *Hpa*II but is much larger than the expected size for the zmet2a insertion. This indicates that this site is methylated in most SPRITE-1 copies.

Another pair of isoschizomers *Bst*NI and *Eco*RII recognize the sequence CC(A/T)GG. *BstN*I is not sensitive to methylation and *Eco*RII will not cut when the internal cytosine is methylated. *Bst*NI should generate SPRITE-1-specific fragments of 6, 54, 135, 252, and 784 bp with the UTR probe. All *Eco*RII fragments were greater than 20 Kb indicating complete methylation of these sites. *Hha*I which recognizes GCGC sites should generate SPRITE-1-specific fragments of 2884 and 257 bp and a zmet2a insertion fragment of 2965 bp. No fragments this small were

observed indicating methylation at these sites. The *Pst*I site recognized with this probe was also methylated.

EXAMPLE 2 – Cloning and Sequencing of zmet2b

A lambda library (Stratagene) constructed from Mo17 maize genomic DNA library was screened with the zmet2a methyltransferase nucleic sequences shown in FIG. 1. This screening resulted in the recovery of seven (7) independent clones. Four of these clones corresponded exactly to zmet2a nucleic acid sequence. Another type, represented by only one clone, had limited homology in non-significant regions. Two other clones were very similar to the zmet2a methyltransferase nucleic acid sequence but were definitely not identical to the zmet2a methyltransferase nucleic acid sequence. These clones defined a second gene, referred to as "zmet2b". Primer walking resulted in a partial genomic sequence of zmet2b. Primers specific to zmet2b were designed and used to amplify zmet2b cDNA (using Marathon cDNA Amplification Kit from Clontech according to the manufacturer's protocols). The RACE products were isolated and cloned into p-GEMT-Easy (Promega). Sequence of the RACE products generated a partial cDNA sequence for the 3' end of the gene (see FIG. 23). A partial amino acid sequence encoded by this cDNA sequence is shown in FIG. 24. A comparison of a portion of the amino acid sequences for zmet2a and zmet2b is shown in FIG. 25.

All references cited herein are hereby incorporated by reference.

The present invention is illustrated by way of the foregoing description and examples. The foregoing description is intended as a non-limiting illustration, since many variations will become apparent to those skilled in the art in view thereof. It is intended that all such variations within the scope and spirit of the appended claims be embraced thereby.

Changes can be made to the composition, operation and arrangement of the method of the present invention described herein without departing from the concept and scope of the invention as defined in the following claims.

WHAT IS CLAIMED IS:

1. An isolated and purified DNA sequence which encodes a *Zea mays* zmet2a methyltransferase and which hybridizes to the nucleic acid sequence shown in FIG. 1A under stringent conditions.

2. An isolated and purified zmet2a methyltransferase comprising the amino acid sequence shown in FIG. 2A.

3. An isolated and purified DNA sequence which encodes a *Zea mays* zmet2b methyltransferase and which hybridizes to the nucleic acid sequence shown in FIG. 1B under stringent conditions.

4. An isolated and purified zmet2a methyltransferase comprising the amino acid sequence shown in FIG. 2B.

5. A recombinant expression cassette comprising the isolated and purified nucleic acid sequence of claims 1 or 3, a promoter sequence and a polyadenylation signal sequence, wherein the promoter sequence is operably linked to the nucleic acid sequence and the nucleic acid sequence is operably linked to the polyadenylation signal sequence.

6. The recombinant expression cassette of claim 5 wherein the promoter sequence is a constitutive or a tissue specific promoter sequence.

7. A bacterial cell comprising the recombinant expression cassette of claim 5.

8. The bacterial cell of claim 7 wherein the bacterial cell is selected from the group consisting of *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*.

9. A transgenic plant comprising the recombinant expression cassette of claim 5.

10. The transgenic plant of claim 9 wherein the promoter sequence and the polyadenylation signal sequence is from Cauliflower Mosaic Virus 35S gene.

11. The transgenic plant of claim 10 wherein transgenic plant is *Zea mays*, *Oryza sativa*, *Secale cereale*, *Triticum aestivum*, *Daucus carota*, *Brassica oleracea*, *Cucumis melo*, *Cucumis sativus*, *Latuca sativa*, *Solanum tubersoum*, *Lycopersicon esculentum*, *Phaseolus vulgaris*, and *Brassica napus*.

12. Seed comprising the recombinant expression cassette of claim 5.

13. An isolated and purified DNA sequence which encodes a *Zea mays* zmet2b methyltransferase and which hybridizes to the nucleic acid sequence of FIG. 23 under stringent conditions.

14. A recombinant expression cassette comprising the isolated and purified nucleic acid sequence of claim 13, a promoter sequence and a polyadenylation signal sequence, wherein the promoter sequence is operably linked to the nucleic acid sequence and the nucleic acid sequence is operably linked to the polyadenylation signal sequence.

15. The recombinant expression cassette of claim 14 wherein the promoter sequence is a constitutive or a tissue specific promoter sequence.

16. A bacterial cell comprising the recombinant expression cassette of claim 14.

17. The bacterial cell of claim 16 wherein the bacterial cell is selected from the group consisting of *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*.

18. A transgenic plant comprising the recombinant expression cassette of claim 14.

19. The transgenic plant of claim 18 wherein the promoter sequence and the polyadenylation signal sequence is from Cauliflower Mosaic Virus 35S gene.

20. The transgenic plant of claim 19 wherein transgenic plant is *Zea mays*, *Oryza sativa*, *Secale cereale*, *Triticum aestivum*, *Daucus carota*, *Brassica oleracea*, *Cucumis melo*, *Cucumis sativus*, *Latuca sativa*, *Solanum tubersoum*, *Lycopersicon esculentum*, *Phaseolus vulgaris*, and *Brassica napus*.

21. Seed comprising the recombinant expression cassette of claim 14.

FIG. 1A 2736 bp

```
   1  ATGGCGCCGA GCTCCCCGTC ACCCGCCGCG CCTACACGCG TCTCTGGGCG
  51  GAAGCGCGCC GCCAAGGCCG AGGAGATCCA CCAGAACAAG GAGGAGGAGG
 101  AGGAGGTCGC GGCGGCGTCC TCCGCCAAGC GCAGCCGCAA GGCGGCATCT
 151  TCCGGGAAGA AGCCCAAGTC GCCCCCCAAG CAGGCCAAGC CGGGGAGGAA
 201  GAAGAAGGGG GATGCCGAGA TGAAGGAGCC CGTGGAGGAC GACGTGTGCG
 251  CCGAGGAGCC CGACGAGGAG GAGTTGGCCA TGGGCGAGGA GGAGGCCGAG
 301  GAGCAGGCCA TGCAGGAGGA GGTGGTTGCG GTCGCGGCGG GGTCACCCGG
 351  GAAGAAGAGG GTGGGGAGAA GGAACGCCGC CGCCGCCGCT GGCGACCACG
 401  AGCCGGAGTT CATCGGCAGC CCTGTTGCCG CGGACGAGGC GCGCAGCAAC
 451  TGGCCCAAGC GCTACGGCCG CAGCACTGCC GCAAAGAAAC CGGATGAGGA
 501  GGAAGAGCTC AAGGCCAGAT GTCACTACCG GAGCGCTAAG GTGGACAACG
 551  TCGTCTACTG CCTCGGGGAT GACGTCTATG TCAAGGCTGG AGAAAACGAG
 601  GCAGATTACA TTGGCCGCAT TACTGAATTT TTTGAGGGGA CTGACCAGTG
 651  TCACTATTTT ACTTGCCGTT GGTTCTTCCG AGCAGAGGAC ACGGTTATCA
 701  ATTCTTTGGT GTCCATAAGT GTGGATGGCC ACAAGCATGA CCCTAGACGT
 751  GTTTTTCTTT CTGAGGAAAA GAACGACAAT GTGCTTGATT GCATTATCTC
 801  CAAGGTCAAG ATAGTCCATG TTGATCCAAA TATGGATCCA AAAGCCAAGG
 851  CTCAGCTGAT AGAGAGTTGC GACCTATACT ATGACATGTC TTACTCTGTT
 901  GCATATTCTA CATTTGCTAA TATCTCGTCT GAAAATGGGC AGTCAGGCAG
 951  TGATACCGCT TCGGGTATTT CTTCTGATGA TGTGGATCTG GAGACGTCAT
1001  CTAGTATGCC AACGAGGACA GCAACCCTTC TTGATCTGTA TTCTGGCTGT
1051  GGGGGCATGT CTACTGGTCT TTGCTTGGGT GCAGCTCTTT CTGGCTTGAA
1101  ACTTGAAACT CGATGGGCTG TTGATTTCAA CAGTTTTGCG TGCCAAAGTT
1151  TAAAATATAA TCATCCACAG ACTGAGGTGC GAAATGAGAA AGCCGATGAG
1201  TTTCTTGCCC TCCTTAAGGA ATGGGCAGTT CTATGCAAAA AATATGTCCA
1251  AGATGTGGAT TCAAATTTAG CAAGCTCAGA GGATCAAGCG GATGAAGACA
1301  GCCCTCTTGA CAAGGACGAA TTTGTTGTAG AGAAGCTTGT CGGGATATGT
1351  TATGGTGGCA GTGACAGGGA AAATGGCATC TATTTTAAGG TCCAGTGGGA
1401  AGGATACGGC CCTGAGGAGG ATACATGGGA ACCGATTGAT AACTTGAGTG
1451  ACTGCCCGCA GAAAATTAGA GAATTTGTAC AAGAAGGGCA CAAAAGAAAG
1501  ATTCTCCCAC TGCCTGGTGA TGTTGATGTC ATTTGTGGAG GCCCACCATG
1551  CCAAGGTATC AGTGGGTTTA ATCGGTACAG AAACCGTGAT GAGCCACTCA
1601  AAGATGAGAA AAACAAACAA ATGGTGACTT TCATGGATAT TGTGGCGTAC
1651  TTGAAGCCCA AGTATGTTCT CATGGAAAAT GTGGTGGACA TACTCAAATT
1701  TGCGGATGGT TACCTAGGAA AATATGCTTT GAGCTGCCTT GTTGCTATGA
1751  AGTACCAAGC GCGGCTTGGA ATGATGGTGG CTGGTTGCTA TGGTCTGCCA
1801  CAGTTCAGGA TGCGTGTGTT CCTCTGGGGT GCTCTTTCTT CCATGGTGCT
1851  CCCTAAGTAT CCTCTGCCCA CCTATGATGT TGTAGTACGT GGAGGAGCCC
1901  CTAATGCCTT TTCGCAATGT ATGGTTGCAT ATGACGAGAC ACAAAAACCA
1951  TCCCTGAAAA AAGCCTTGCT TCTTGGCGAT GCAATTTCAG ATTTACCAAA
2001  GGTTCAAAAT CACCAGCCTA ACGATGTGAT GGAGTATGGT GGTTCCCCCA
2051  AGACCGAATT CCAGCGCTAC ATTCGACTCA GTCGTAAAGA CATGTTGGAT
2101  TGGTCCTTCG GTGAGGGGGC TGGTCCAGAT GAAGGCAAGC TCTTGGATCA
2151  CCAGCCTTTA CGGCTTAACA ACGATGATTA TGAGCGGGTT CAACAGATTC
2201  CTGTCAAGAA GGGAGCCAAC TTCCGCGACC TAAAGGGCGT GAGGGTTGGA
2251  GCAAACAATA TTGTTGAGTG GGATCCAGAA ATCGAGCGTG TGAAACTTTC
2301  ATCTGGGAAA CCACTGGTTC CTGACTATGC AATGTCATTC ATCAAGGGCA
2351  AATCACTCAA GCCGTTTGGG CGCCTGTGGT GGGACGAGAC AGTTCCTACA
2401  GTTGTAACCA GAGCAGAGCC TCACAACCAG GTTATAATTC ATCCGACTCA
2451  AGCAAGGGTC CTCACTATCC GGGAGAACGC AAGGTTACAG GGTTCCCCG
2501  ATTACTACCG ATTGTTTGGC CCGATCAAGG AGAAGTACAT TCAAGTCGGG
2551  AACGCAGTGG CTGTCCCTGT TGCCCGGCA CTGGGCTACT GTCTGGGGCA
2601  AGCCTACCTG GGTGAATCTG AGGGGAGTGA CCCTCTGTAC CAGCTGCCTC
2651  CAAGTTTCAC CTCTGTTGGA GGACGCACTG CGGGGCAGGC GAGGGCCTCT
2701  CCTGTTGGCA CCCCTGCAGG GGAGGTAGTT GAGCAG
```

FIG. 1B

```
   1  AGAGCAGCAG CAGCTACCGC AGCCCCTGCC ATGGCGCCGA GCTCCCCGTC
  51  ACCCGCCGCG CCTACACGCG TCTCTGGGCG GAAGCGCGCC GCCAAGGCCG
 101  AGGAGATCCA CCAGAACAAG GAGGAGGAGG AGGAGGTCGC GGCGGCGTCC
 151  TCCGCCAAGC GCAGCCGCAA GGCGGCATCT TCCGGGAAGA AGCCCAAGTC
 201  GCCCCCCAAG CAGGCCAAGC CGGGGAGGAA GAAGAAGGGG GATGCCGAGA
 251  TGAAGGAGCC CGTGGAGGAC GACGTGTGCG CCGAGGAGCC CGACGAGGAG
 301  GAGTTGGCCA TGGGCCAGGA GGAGGCCGAG GAGCAGGCCA TGCAGGAGGA
 351  GGTGGTTGCG GTCGCGGCGG GGTCACCCGG GAAGAAGAGG GTGGGGAGAA
 401  GGAACGCCGC CGCCGCCGCT GGCGACCACG AGCCGGAGTT CATCGGCAGC
 451  CCTGTTGCCG CGGACGAGGC GCGCAGCAAC TGGCCCAAGC GCTACGGCCG
 501  CAGCACTGCC GCAAAGAAAC CGGATGAGGA GGAAGAGCTC AAGGCCAGAT
 551  GTCACTACCG GAGCGCTAAG GTGGACAACG TCGTCTACTG CCTCGGGGAT
 601  GACGTCTATG TCAAGGCTGG AGAAAACGAG GCAGATTACA TTGGCCGCAT
 651  TACTGAATTT TTTGAGGGGA CTGACCAGTG TCACTATTTT ACTTGCCGTT
 701  GGTTCTTCCG AGCAGAGGAC ACGGTTATCA ATTCTTTGGT GTCCATAAGT
 751  GTGGATGGCC ACAAGCATGA CCCTAGACGT GTTTTTCTTT CTGAGGAAAA
 801  GAACGACAAT GTGCTTGATT GCATTATCTC CAAGGTCAAG ATAGTCCATG
 851  TTGATCCAAA TATGGATCCA AAAGCCAAGG CTCAGCTGAT AGAGAGTTGC
 901  GACCTATACT ATGACATGTC TTACTCTGTT GCATATTCTA CATTTGCTAA
 951  TATCTCGTCT GAAAATGGGC AGTCAGGCAG TGATACCGCT TCGGGTATTT
1001  CTTCTGATGA TGTGGATCTG GAGACGTCAT CTAGTATGCC AACGAGGACA
1051  GCAACCCTTC TTGATCTGTA TTCTGGCTGT GGGGGCATGT CTACTGGTCT
1101  TTGCTTGGGT GCAGCTCTTT CTGGCTTGAA ACTTGAAACT CGATGGGCTG
1151  TTGATTTCAA CAGTTTTGCG TGCCAAAGTT TAAAATATAA TCATCCACAG
1201  ACTGAGGTGC GAAATGAGAA AGCCGATGAG TTTCTTGCCC TCCTTAAGGA
1251  ATGGGCAGTT CTATGCAAAA AATATGTCCA AGATGTGGAT TCAAATTTAG
1301  CAAGCTCAGA GGATCAAGCG GATGAAGACA GCCCTCTTGA CAAGGACGAA
1351  TTTGTTGTAG AGAAGCTTGT CGGGATATGT TATGGTGGCA GTGACAGGGA
1401  AAATGGCATC TATTTTAAGG TCCAGTGGGA AGGATACGGC CCTGAGGAGG
1451  ATACATGGGA ACCGATTGAT AACTTGAGTG ACTGCCCGCA GAAAATTAGA
1501  GAATTTGTAC AAGAAGGGCA CAAAAGAAAG ATTCTCCCAC TGCCTGGTGA
1551  TGTTGATGTC ATTTGTGGAG GCCCACCATG CCAAGGTATC AGTGGGTTTA
1601  ATCGGTACAG AAACCGTGAT GAGCCACTCA AAGATGAGAA AAACAAACAA
1651  ATGGTGACTT TCATGGATAT TGTGGCGTAC TTGAAGCCCA AGTATGTTCT
1701  CATGGAAAAT GTGGTGGACA TACTCAAATT TGCGGATGGT TACCTAGGAA
1751  AATATGCTTT GAGCTGCCTT GTTGCTATGA AGTACCAAGC GCGGCTTGGA
1801  ATGATGGTGG CTGGTTGCTA TGGTCTGCCA CAGTTCAGGA TGCGTGTGTT
1851  CCTCTGGGGT GCTCTTTCTT CCATGGTGCT CCCTAAGTAT CCTCTGCCCA
1901  CCTATGATGT TGTAGTACGT GGAGGAGCCC CTAATGCCTT TTCGCAATGT
1951  ATGGTTGCAT ATGACGAGAC ACAAAAACCA TCCCTGAAAA AAGCCTTGCT
2001  TCTTGGCGAT GCAATTTCAG ATTTACCAAA GGTTCAAAAT CACCAGCCTA
2051  ACGATGTGAT GGAGTATGGT GGTTCCCCCA AGACCGAATT CCAGCGCTAC
2101  ATTCGACTCA GTCGTAAAGA CATGTTGGAT TGGTCCTTCG GTGAGGGGGC
2151  TGGTCCAGAT GAAGGCAAGC TCTTGGATCA CCAGCCTTTA CGGCTTAACA
2201  ACGATGATTA TGAGCGGGTT CAACAGATTC CTGTCAAGAA GGGAGCCAAC
2251  TTCCGCGACC TAAAGGGCGT GAGGGTTGGA GCAAACAATA TTGTTGAGTG
2301  GGATCCAGAA ATCGAGCGTG TGAAACTTTC ATCTGGGAAA CCACTGGTTC
2351  CTGACTATGC AATGTCATTC ATCAAGGGCA AATCACTCAA GCCGTTTGGG
2401  CGCCTGTGGT GGGACGAGAC AGTTCCTACA GTTGTAACCA GAGCAGAGCC
2451  TCACAACCAG GTTATAATTC ATCCGACTCA AGCAAGGGTC CTCACTATCC
2501  GGGAGAACGC AAGGTTACAG GGCTTCCCCG ATTACTACCG ATTGTTTGGC
2551  CCGATCAAGG AGAAGTACAT TCAAGTCGGG AACGCAGTGG CTGTCCCTGT
2601  TGCCCGGGCA CTGGGCTACT GTCTGGGGCA AGCCTACCTG GGTGAATCTG
2651  AGGGGAGTGA CCCTCTGTAC CAGCTGCCTC CAAGTTTCAC CTCTGTTGGA
2701  GGACGCACTG CGGGGCAGGC GAGGGCCTCT CCTGTTGGCA CCCCTGCAGG
2751  GGAGGTAGTT GAGCAGTAAA AGGATGACAG ATCTGAGCTG AGCTGG
```

FIG. 2A

912 amino acids

```
   1  MAPSSPSPAA PTRVSGRKRA AKAEEIHQNK EEEEEVAAAS SAKRSRKAAS
  51  SGKKPKSPPK QAKPGRKKKG CAEMKEPVED DVCAEEPDEE ELAMGEEEAE
 101  EQAMQEEVVA VAAGSPGKKR VGRRNAAAAA GDHEPEFIGS PVAADEARSN
 151  WPKRYGRSTA AKKPDEEEEL KARCHYRSAK VDNVVYCLGD DVYVKAGENE
 201  ADYIGRITEF FEGTDQCHYF TCRWFFRAED TVINSLVSIS VDGHKHDPRR
 251  VFLSEEKNDN VLDCIISKVK IVHVDPNMDP KAKAQLIESC DLYYDMSYSV
 301  AYSTFANISS ENGQSGSDTA SGISSDDVDL ETSSSMPTRT ATLLDLYSGC
 351  GGMSTGLCLG AALSGLKLET RWAVDFNSFA CQSLKYNHPQ TEVRNEKADE
 401  FLALLKEWAV LCKKYVQDVD SNLASSEDQA DEDSPLDKDE FVVEKLVGIC
 451  YGGSDRENGI YFKVQWEGYG PEEDTWEPID NLSDCPQKIR EFVQEGHKRK
 501  ILPLPGDVDV ICGGPPCQGI SGFNRYRNRD EPLKDEKNKQ MVTFMDIVAY
 551  LKPKYVLMEN VVDILKFADG YLGKYALSCL VAMKYQARLG MMVAGCYGLP
 601  QFRMRVFLWG ALSSMVLPKY PLPTYDVVVR GGAPNAFSQC MVAYDETQKP
 651  SLKKALLLGD AISDLPKVQN HQPNDVMEYG GSPKTEFQRY IRLSRKDMLD
 701  WSFGEGAGPD EGKLLDHQPL RLNNDDYERV QQIPVKKGAN FRDLKGVRVG
 751  ANNIVEWDPE IERVKLSSGK PLVPDYAMSF IKGKSLKPFG RLWWDETVPT
 801  VVTRAEPHNQ VIIHPTQARV LTIRENARLQ GFPDYYRLFG PIKEKYIQVG
 851  NAVAVPVARA LGYCLGQAYL GESEGSDPLY QLPPSFTSVG GRTAGQARAS
 901  PVGTPAGEVV EQ
```

FIG. 2B

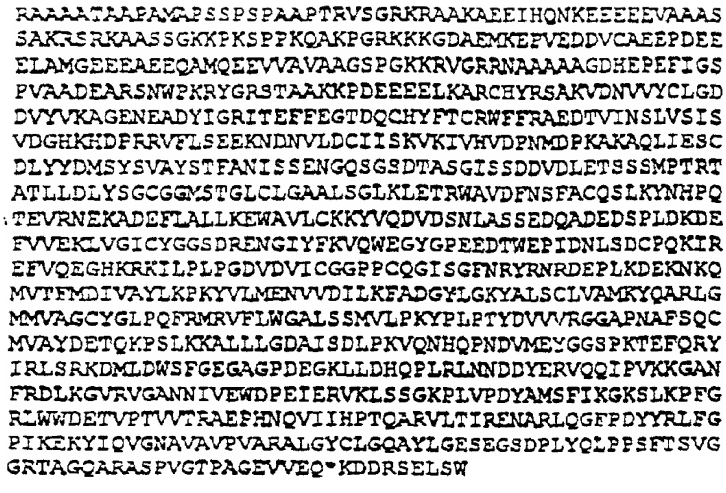

RAAAATAAPAMAPSSPSPAAPTRVSGRKRAAKAEEIHQNKEEEEEVAAAS
SAKRSRKAASSGKKPKSPPKQAKPGRKKKGDAEMKEPVEDDVCAEEPDEE
ELAMGEEEAEEQAMQEEVVAVAAGSPGKKRVGRRNAAAAAGDHEPEFIGS
PVAADEARSNWPKRYGRSTAAKKPDEEEELKARCHYRSAKVDNVVYCLGD
DVYVKAGENEADYIGRITEFFEGTDQCHYFTCRWFFRAEDTVINSLVSIS
VDGHKHDFRRVFLSEEKNDNVLDCIISKVKIVHVDPNMDPKAKAQLIESC
DLYYDMSYSVAYSTFANISSENGQSGSDTASGISSDDVDLETSSSMPTRT
ATLLDLYSGCGGMSTGLCLGAALSGLKLETRWAVDFNSFACQSLKYNHPQ
TEVRNEKADEFLALLKEWAVLCKKYVQDVDSNLASSEDQADEDSPLDKDE
FVVEKLVGICYGGSDRENGIYFKVQWEGYGPEEDTWEPIDNLSDCPQKIR
EFVQEGHKRKILPLPGDVDVICGGPPCQGISGFNRYRNRDEPLKDEKNKQ
MVTFMDIVAYLKPKYVLMENVVDILKFADGYLGKYALSCLVAMKYQARLG
MMVAGCYGLPQFRMRVFLWGALSSMVLPKYPLPTYDVVVRGGAPNAFSQC
MVAYDETQKPSLKKALLLGDAISDLPKVQNHQPNDVMEYGGSPKTEFQRY
IRLSRKDMLDWSFGEGAGPDEGKLLDHQPLRLNNDDYERVQQIPVKKGAN
FRDLKGVRVGANNIVEWDPEIERVKLSSGKPLVPDYAMSFIKGKSLKPFG
RLWWDETVPTVVTRAEPHNQVIIHPTQARVLTIRENARLQGFPDYYRLFG
PIKEKYIQVGNAVAVPVARALGYCLGQAYLGESEGSDPLYQLPPSFTSVG
GRTAGQARASPVGTPAGEVVEQ*KDDRSELSW

FIG. 3

Primer	Sequence 5' - 3'
1F	TGGTTGCTATGGTCTGCCACAGTTCAG
1R	CCAGCTCAGCTCAGATCTGTCATCCTTT
Seq2FN	CGAAAGCTAATCTACACAAACAGC
Seq2RN	GATCCTCTGAGCTTGCTAAATTTG
3R	CTCATCTTGGAGTGGCTCATCAC
S3F	GAGCACATGAGGGAGAGTGTTG
S3R	TCTCTAATTTTCTGCGGGCAG
4F	CCTCTGCCCACCTATGATGTTGTA
5F	TAAAGGGCGTGAGGGTTGGA
7F	TCACATTTGTCATGGCAGGTTATC
8eF	CTGAGGAAAAGAACGACAATGTGC
8eR	GCAATCAAGCACATTGTCGTTCTTTTCCTC
9eF	GAAGAAGAGGGTGGGGAGAAGGAACG
9eR	TTCTTTGCGGCAGTGCTGCG
11iF	GTATTGAATTGATTCTCAACTAGTGCAC
11iR	CAGGCTCAACGGCGATG
12iF	TATGCTTCATCACATAGACCCAAGTC
12iR	GATAGACCTAATGCCAAATGAGATTAAG
13iF	GCGATCTTCAGTCTCCACCATC
13iR	GAAGACGTGCCTCCATGTTTCATC
14F	GTTGGTTCTTCCGAGCAGAGG
14R	GACTGCCACATATCTTATTAATCGC
15F	GCATGTGTCAGCAATTGCTTACATTC
15R	CCTCTGCTCGGAAGAACCAAC
16F	CTGTTCGGAGATTCATGCATGATG
16R	GGAGAACAGAATGGTTGATTCAATGG
17F	GCACTTCACTCTCCTGGCAAACC
17R	CGGTACGCTGCTGCTGCTCTC
18F	CCATAGCATCTCACATATCGCAAGG
18R	GGAAAGAAGGCAGTTAGTTGTAAATGGG
MU	AGAGAAGCCAACGCCAWCGCCTCYATTTCGTC
RaceRT	CTACAACATCATAGTTGGGCAGAGG
AP2 marathon	ACTCACTATAGGGCTCGAGCGGC
T7	TAATACGACTCACTATAGGG
Sp6	GATTTAGGTGACACTATAG
M13F	GTTTTCCCAGTCACGAC
M13R	CAGGAAACAGCTATGAC

FIG. 4

Gene Name	Organism	Function	General Structure
DNMT1/Dnmt1	human/mouse	maintenance	N-terminal domain; Methylase domain; I IV VI VIII IX X
MET1	*Arabidopsis*	maintenance	NLS; Replication foci Target sequence; Lys-Gly repeat
Zmet1	maize	putative maintenance	
DNMT3/Dnmt3	human/mouse	*de novo*	Cys-rich region
Zmet3	maize	putative *de novo*	
DRM	*Arabidopsis*	putative *de novo*	
CMT1	*Arabidopsis*	undetermined (putative CpNpG)	chromodomains
Zmet2a,	maize	CpNpG (maintenance and/or *de novo*)	

Figure 5

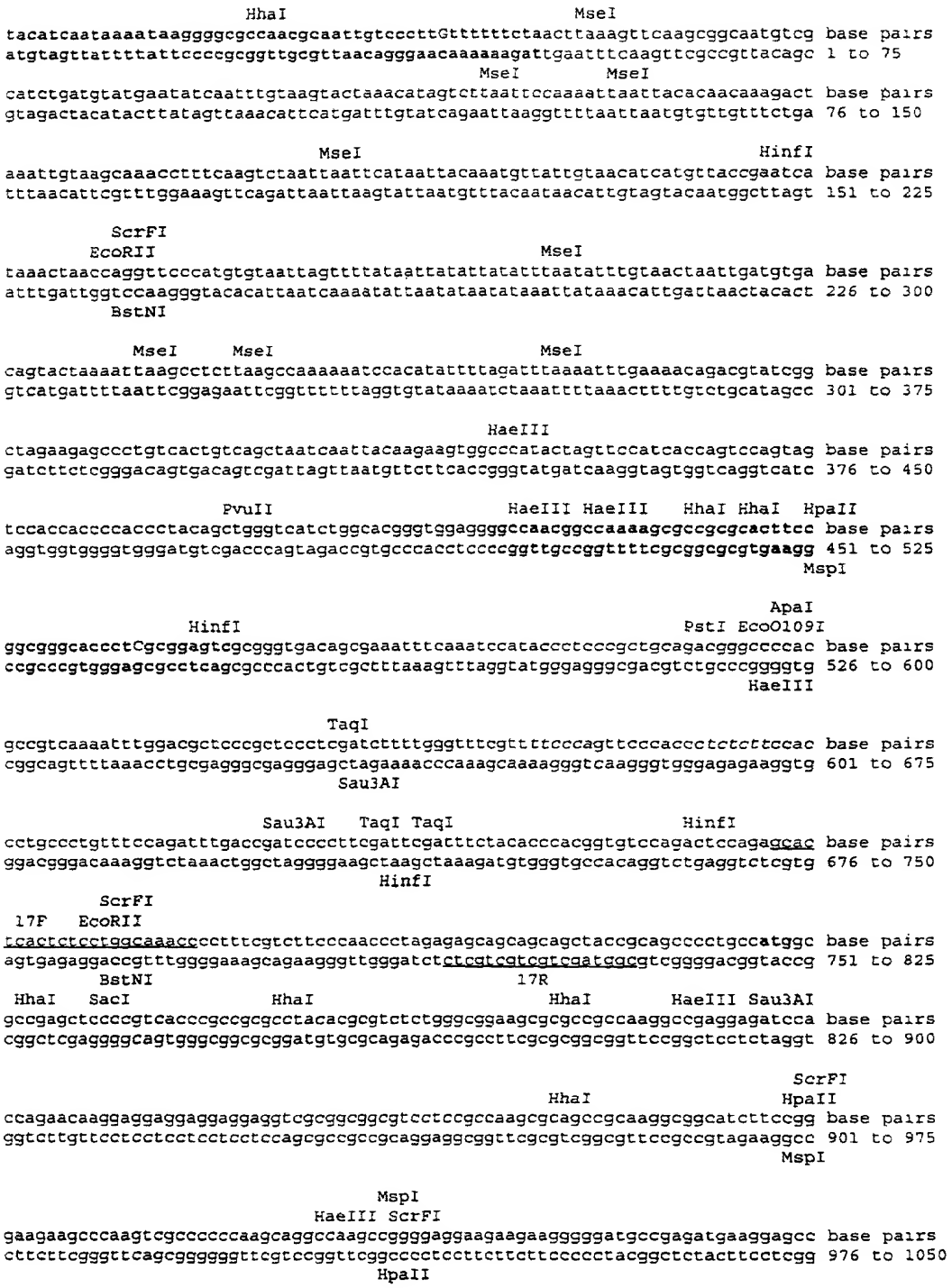

HhaI MseI
tacatcaataaaataaggggcgccaacgcaattgtcccttGttttttctaacttaaagttcaagcggcaatgtcg base pairs
atgtagttatttttattccccgcggttgcgttaacagggaacaaaaaagattgaatttcaagttcgccgttacagc 1 to 75
MseI MseI
catctgatgtatgaatatcaatttgtaagtactaaacatagtcttaattccaaaattaattacacaacaaagact base pairs
gtagactacatacttatagttaaacattcatgatttgtatcagaattaaggttttaattaatgtgttgtttctga 76 to 150
MseI HinfI
aaattgtaagcaaacctttcaagtctaattaattcataattacaaatgttattgtaacatcatgttaccgaatca base pairs
tttaacattcgtttggaaagttcagattaattaagtattaatgtttacaataacattgtagtacaatggcttagt 151 to 225
ScrFI
EcoRII MseI
taaactaaccaggttcccatgtgtaattagttttataattatattatatttaatatttgtaactaattgatgtga base pairs
atttgattggtccaagggtacacattaatcaaaatattaatataatataaattataaacattgattaactacact 226 to 300
BstNI
MseI MseI MseI
cagtactaaaattaagcctcttaagccaaaaaatccacatattttagatttaaaatttgaaaacagacgtatcgg base pairs
gtcatgattttaattcggagaattcggtttttttaggtgtataaaatctaaattttaaacttttgtctgcatagcc 301 to 375
HaeIII
ctagaagagccctgtcactgtcagctaatcaattacaagaagtggcccatactagttccatcaccagtccagtag base pairs
gatcttctcgggacagtgacagtcgattagttaatgttcttcaccgggtatgatcaaggtagtggtcaggtcatc 376 to 450
PvuII HaeIII HaeIII HhaI HhaI HpaII
tccaccaccccaccctacagctgggtcatctggcacgggtggaggggccaacggccaaaagcgccgcgcacttcc base pairs
aggtggtggggtgggatgtcgacccagtagaccgtgcccacctccccggttgccggttttcgcggcgcgtgaagg 451 to 525
MspI
ApaI
HinfI PstI EcoO109I
ggcgggcaccctCgcggagtcgcgggtgacagcgaaatttcaaatccataccctcccgctgcagacgggccccac base pairs
ccgcccgtgggagcgcctcagcgcccactgtcgctttaaagtttaggtatgggagggcgacgtctgcccggggtg 526 to 600
HaeIII
TaqI
gccgtcaaaatttggacgctcccgctccctcgatcttttgggtttcgttttcccagttcccaccctctcttccac base pairs
cggcagttttaaacctgcgagggcgagggagctagaaaacccaaagcaaaagggtcaagggtgggagagaaggtg 601 to 675
Sau3AI
Sau3AI TaqI TaqI HinfI
cctgccctgtttccagatttgaccgatccccttcgattcgatttctacacccacggtgtccagactccagaggcac base pairs
ggacgggacaaaggtctaaactggctaggggaagctaagctaaagatgtgggtgccacaggtctgaggtctcgtg 676 to 750
HinfI
ScrFI
17F EcoRII
tcactctcctggcaaacccctttcgtcttcccaaccctagagagcagcagcagctaccgcagcccctgccatggc base pairs
agtgagaggaccgtttggggaaagcagaagggttgggatctctcgtcgtcgtcgatggcgtcggggacggtaccg 751 to 825
BstNI 17R
HhaI SacI HhaI HhaI HaeIII Sau3AI
gccgagctccccgtcacccgccgcgcctacacgcgtctctgggcggaagcgcgccgccaaggccgaggagatcca base pairs
cggctcgaggggcagtgggcggcgcggatgtgcgcagagacccgccttcgcgcggcggttccggctcctctaggt 826 to 900
ScrFI
HhaI HpaII
ccagaacaaggaggaggaggaggaggtcgcggcggcgtcctccgccaagcgcagccgcaaggcggcatcttccgg base pairs
ggtcttgttcctcctcctcctcctccagcgccgccgcaggaggcggttcgcgtcggcgttccgccgtagaaggcc 901 to 975
MspI
MspI
HaeIII ScrFI
gaagaagcccaagtcgccccccaagcaggccaagccggggaggaagaagaaggggatgccgagatgaaggagcc base pairs
cttcttcgggttcagcgggggttcgtccggttcggccccctccttcttcttcccccctacggctctacttcctcgg 976 to 1050
HpaII

FIG. 5 Continued

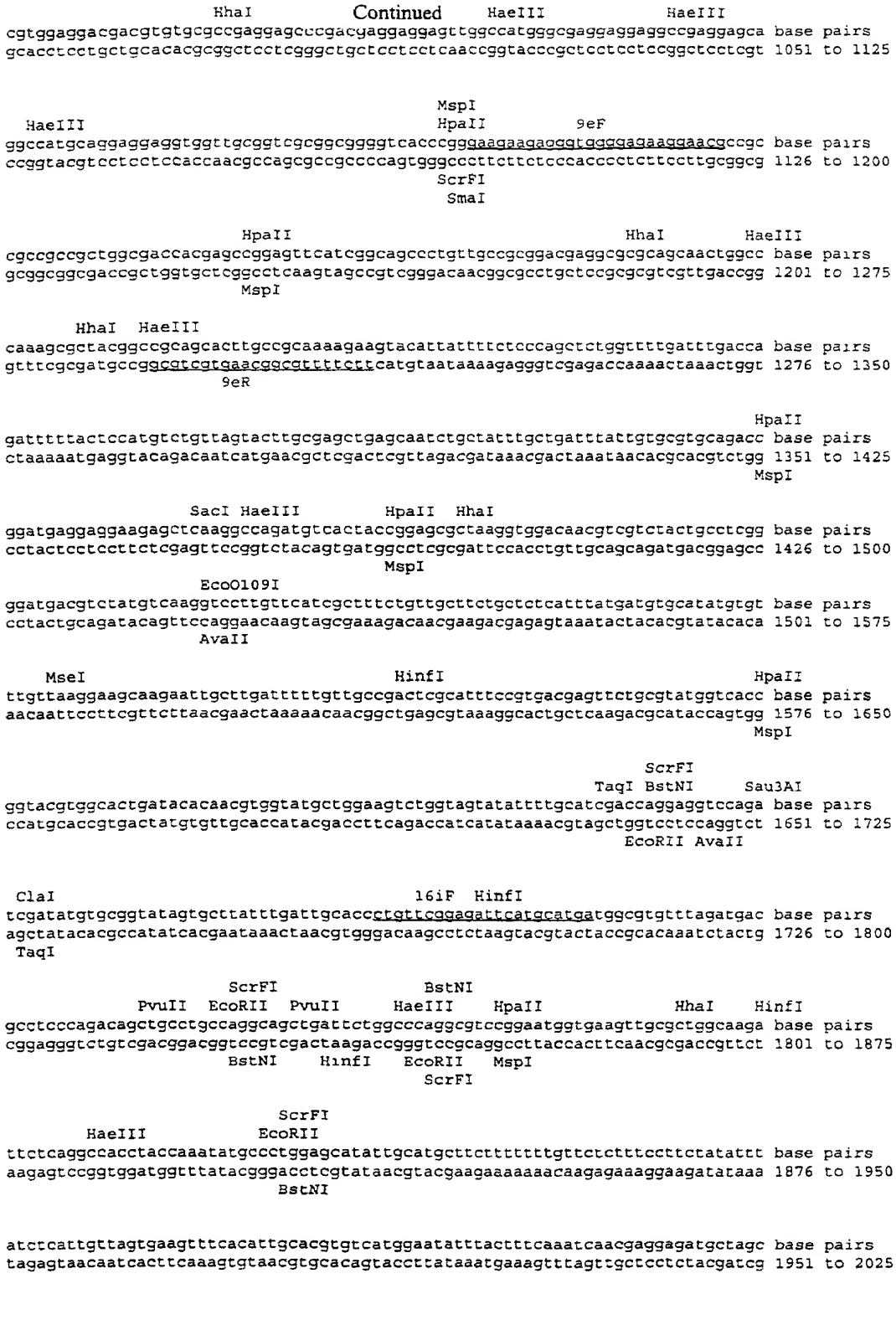


FIG. 5
Continued

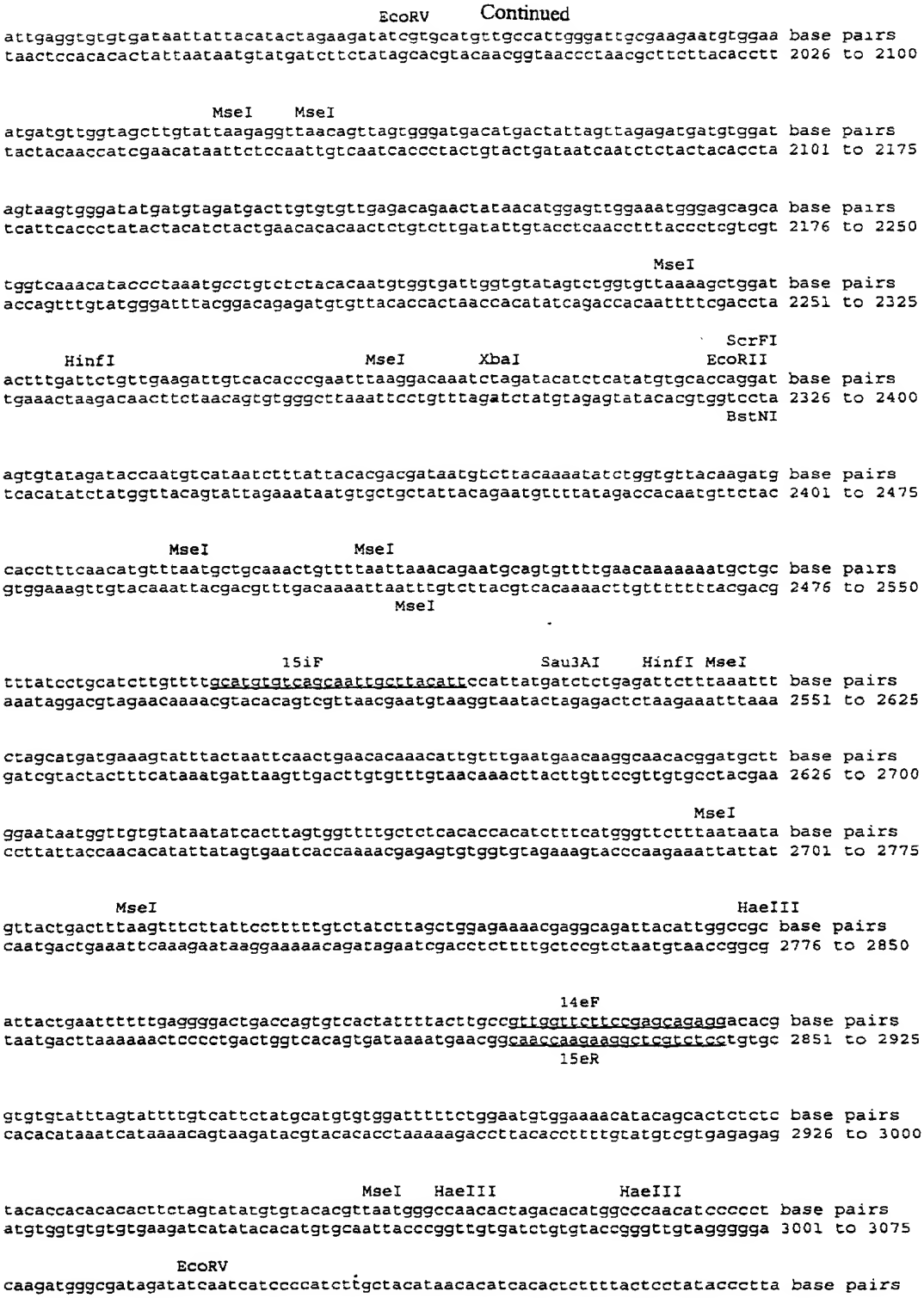


FIG. 5
Continued

```
gttctacccgctatctatagttagtaggggtagaacgatgtattgtgtagtgtgagaaaatgaggatatgggaat 3076 to 3150

                                    HinfI
gtcaagcaatctgctatttgacctttgagtttacatgattcaactctaaagtaccattatctaacttctctttg base pairs
cagttcgttagacgataaactggaaaactcaaatgtactaagttgagatttcatggtaatagattgaagagaaac 3151 to 3225

        ClaI
     HinfI
atgaagaatcgatcaatttccacatgttttgttctatcatgttgaactggattgttagctatattcatggctgac base pairs
tacttcttagctagttaaaggtgtacaaaacaagatagtacaacttgacctaacaatcgatataagtaccgactg 3226 to 3300
        TaqI
         Sau3AI

                          HinfI     MseI                                  18eF
ttattatcacaccataacttcagggagtcttttcttaatacattcaactctgataagagacccttttatccatagc base pairs
aataatagtgtggtattgaagtccctcagaaaagaattatgtaagttgagactattctctgggaaataggtatcg 3301 to 3375

                    HaeIII
atctcacatatcgcaagggccatagctcggtattctgcttcggcggtggaacgggataccacagattgtttcttg base pairs
tagagtgtatagcgttcccggtatcgagccataagacgaagccgccaccttgccctatggtgtctaacaaagaac 3376 to 3450


cttctccatgatactaaatttcctccaacaaacacacaatatcctgaagttgaccttctatcatcaaggcaacta base pairs
gaagaggtactatgatttaaaggaggttgtttgtgtgttataggacttcaactggaagatagtagttccgttgat 3451 to 3525

                                                          ScrFI
                                              MseI        EcoRII
ccccagtctgcatcagagtaaccttccacctttagatgaccatgacctttaaagattattccctttccaggacaa base pairs
ggggtcagacgtagtctcattggaaggtggaaatctactggtactggaaatttctaataagggaaaggtcctgtt 3526 to 3600
                                                          BstNI

                                                        TaqI
gtcttcaagtatcgcagtatacgatacactgcatcaagatgtccacttctggggtcatgcatatatcgactcacc base pairs
cagaagttcatagcgtcatatgctatgtgacgtagttctacaggtgaagaccccagtacgtatatagctgagtgg 3601 to 3675
                                                          HinfI

                    EcoRV
acactgactgcatatgtgatatcaggtcttgtatggcacaagtagatgagccgtccaacaagtctttgatacctt base pairs
tgtgactgacgtatacactatagtccagaacataccgtgttcatctactcggcaggttgttcagaaactatggaa 3676 to 3750

              Sau3AI  HinfI              HinfI  TaqI              HaeIII
tctttattcacaggatcaccagattcagcacataatttatgattcaagtcgataggtgttgctacaggccgacac base pairs
agaaataagtgtcctagtggtctaagtcgtgtattaaatactaagttcagctatccacaacgatgtccggctgtg 3751 to 3825

                        Sau3AI                                        Sau3AI
cccaacatacctgtttcatcaagtagatctaaaacatatttccttgggagagaactattccttttggagatcga base pairs
gggttgtatggacaaagtagttcatctagattttgtataaaggaaaccctctcttgataaggaaaacctctagct 3826 to 3900
                        BglII                                           TaqI

                                    Sau3AI                      HinfI
gcaatctcaataccaagaaagtatttgagatgaccaagatctttaacctcaaattccttacttagattcttcttt base pairs
cgttagagttatggttctttcataaactctactggttctagaaattggagtttaaggaatgaatctaagaagaaa 3901 to 3975
                                    BglII MseI

               Sau3AI
agacatgcaatctcaagatcggcatcacctgtaataataatatcatccacatacacagctagaattgcaattcgt base pairs
tctgtacgttagagttctagccgtagtggacattattattatagtaggtgtatgtgtcgatcttaacgttaagca 3976 to 4050

                        Sau3AI
cgtccaaagtgttgataaaaaacagtgtgatctccgttgcattgtttatatcccatgctacatattgcacgtcta base pairs
gcaggtttcacaactatttttttgtcacactagaggcaacgtaacaaatatagggtacgatgtataacgtgcagat 4051 to 4125

                                                    TaqI
aatctgtcaaaccatgctcttggggactgcttgagaccatacaatgatttttttcaatcgacaaacttttcccaatt base pairs
```

FIG. 5
Continued

```
ttagacagtttggtacgagaacccctgacgaactctggtatgttactaaaaaagttagctgtttgaaagggttaa 4126 to 4200

                    ScrFI
                  EcoRII  Sau3AI
gtctcaggctttgacaatccaggagggatctccatatagacctcctcttgcaaatcaccatgtaagaaagcattc base pairs
cagagtccgaaactgttaggtcctccctagaggtatatctggaggagaacgtttagtggtacattctttcgtaag 4201 to 4275
                    BstNI

 MseI                HaeIII                    Sau3AI
ttaacatctagttgatacaagggccatccaaaatttgcagcacaagagatcaatgtccttacagtactcatttttt base pairs
aattgtagatcaactatgttcccggtaggttttaaacgtcgtgttctctagttacaggaatgtcatgagtaaaaa 4276 to 4350


gccactggtgcaaatgtctcatcataatcaattccatatgtttgactataccctcttgcaaccaatcttgcttta base pairs
cggtgaccacgtttacagagtagtattagttaaggtatacaaactgatatgggagaacgttggttagaacgaaat 4351 to 4425


tatcgttctacccttccttctgggttttgcttcacagtgaatacccatttacaactaactgccttctttccttta base pairs
atagcaagatgggaaggaagacccaaaacgaagtgtcacttatgggtaaatgttgattgacggaagaaaggaaat 4426 to 4500
                                                  181R
                                  XbaI      MseI
ggtagtttctcaaattcccaagtttgatttttttctagagctttaagctcctccaacattgcctcacgccagtta base pairs
ccatcaaagagtttaagggttcaaactaaaaaaagatctcgaaattcgaggaggttgtaacggagtgcggtcaat 4501 to 4575


gaattacattgtgcttctttccaatctcttggaattgctacggaatgcaatgatgcaacaaatgctctatatgat base pairs
cttaatgtaacacgaagaaaggttagagaaccttaacgatgccttacgttactacgttgtttacgagatatacta 4576 to 4650


                                                                    HinfI
ggtgacaaagacgcatatgagacataattgctaatgtcatgttcatatccatacctttgttgggggggactccagct base pairs
ccactgtttctgcgtatactctgtattaacgattacagtacaagtataggtatggaacaacccccctgaggtcga 4651 to 4725

          HhaI
ttagcacgcgctccttttcgtattgcaatgggcaaatcataagtgtcataatcttcagtttctccatgagacgtc base pairs
aatcgtgcgcgaggaaaagcataacgttacccgtttagtattcacagtattagaagtcaaagaggtactctgcag 4726 to 4800


aaaggtacatttatagcctctaatgtgtttggagagaactgctcagtacttgatgctgaattggtttcaggagcc base pairs
tttccatgtaaatatcggagattacacaaacctctcttgacgagtcatgaactacgacttaaccaaagtcctcgg 4801 to 4875

tgaggttgcacatgggacttttcttcttgtatatacttcgcccttatatcgtaagtcgtctccacaagatttatta base pairs
actccaacgtgtaccctgaaagaagaacatatatgaagcgggaatatagcattcagcagaggtgttctaaataat 4876 to 4950


ttctcgtgactaggatgtgtctccaattcacttggcattacttgcatcttttgagaagcaccaatcaccacttcc base pairs
aagagcactgatcctacacagaggttaagtgaaccgtaatgaacgtagaaaactcttcgtggttagtggtgaagg 4951 to 5025

                          HinfI                     TaqI
atttttatttggttgttgttccattgaatcaaccattctgttctccccctctcgactagcttcatctgtgctagta base pairs
taaaataaaccaacaacaaggtaacttagttggtaagacaagaggggggagagctgatcgaagtagacacgatcat 5026 to 5100
                        16iR

       HinfI              Sau3AI
gagacagaatcaagaaaaaaatttagatctgtcttctcaccatagaaaggcacagtctctctaaatgtaacatcc base pairs
ctctgtcttagttctttttttaaatctagacagaagagtggtatctttccgtgtcagagagatttacattgtagg 5101 to 5175
                          BglII

                                HinfI                                PstI
atgcttacaaacaaacgtcgttcactaggactccaacacttgtatccctttgccctgcaggatatccaacaaaa base pairs
tacgaatgtttgtttgcagcaagtgatcctgaggttgtgaacatagggaaaacgggacgtcctataggttgtttt 5176 to 5250
                                                                  EcoRV

                    BamHI                           Sau3AI
atgcacttcacagcacgaggatccaacttccccacctgaggtctatgatctctgacaaaacatgtacatccaaaa base pairs
tacgtgaagtgtcgtgctcctaggttgaaggggtggactccagatactagagactgttttgtacatgtaggtttt 5251 to 5325
```

FIG. 5
Continued

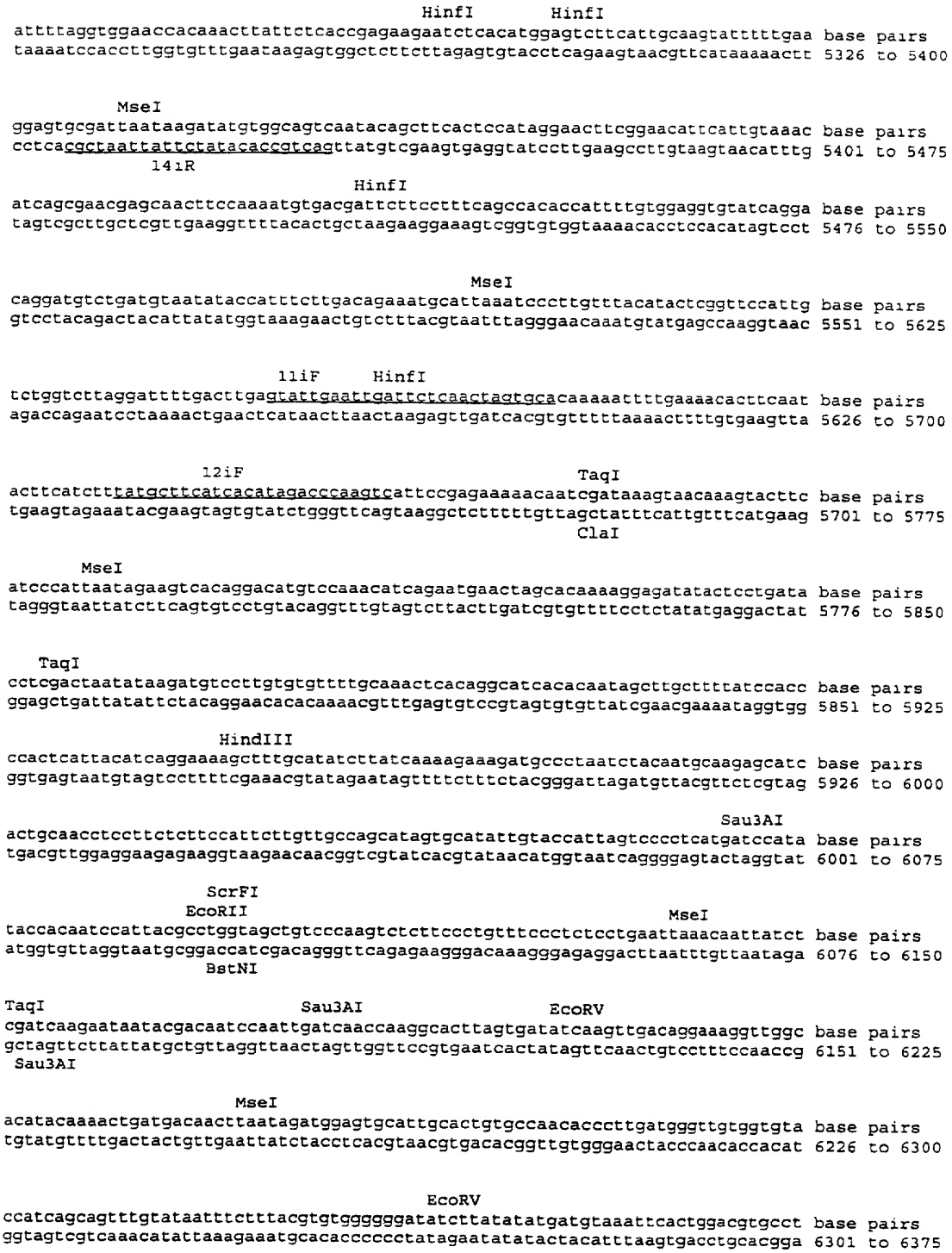
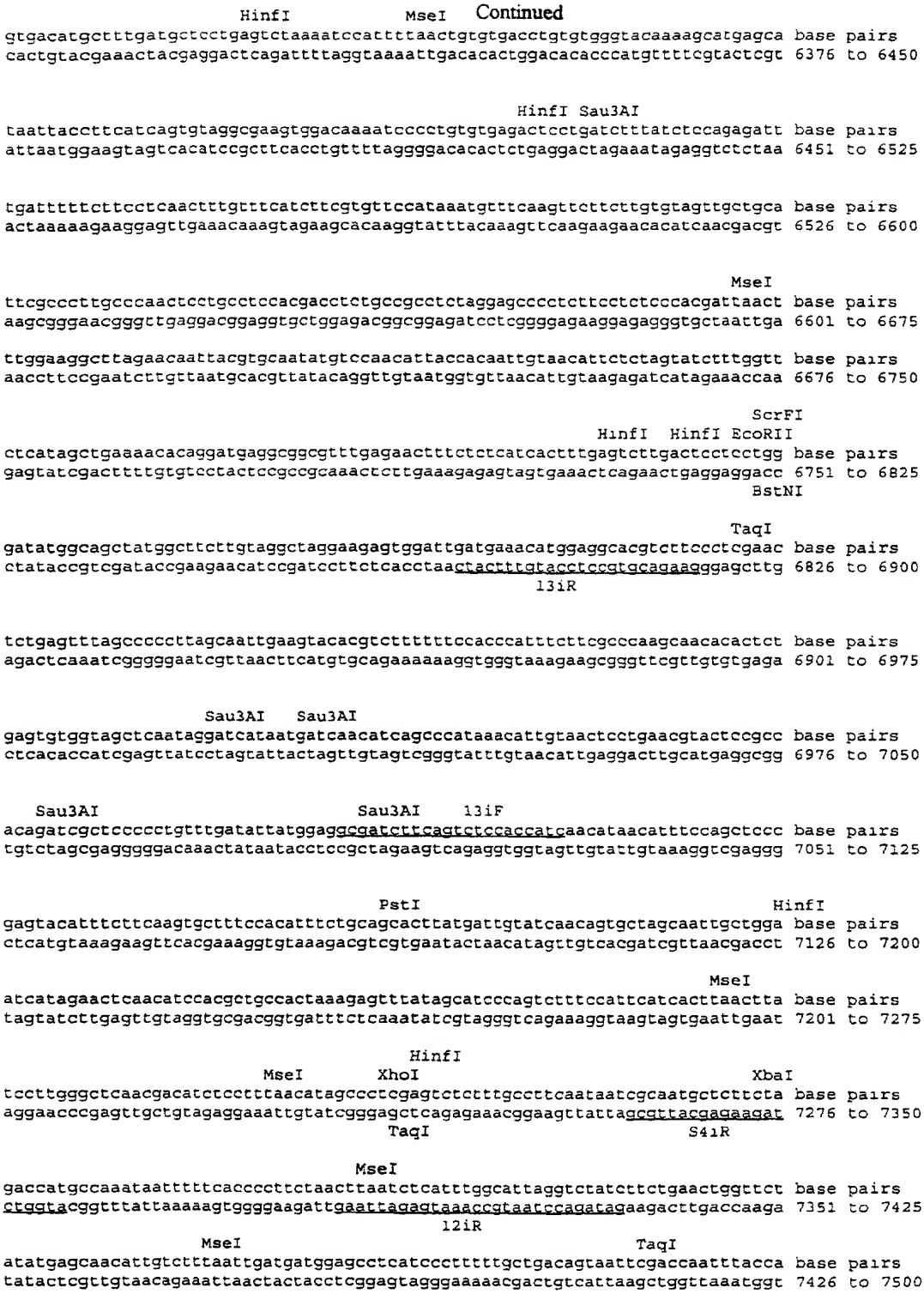


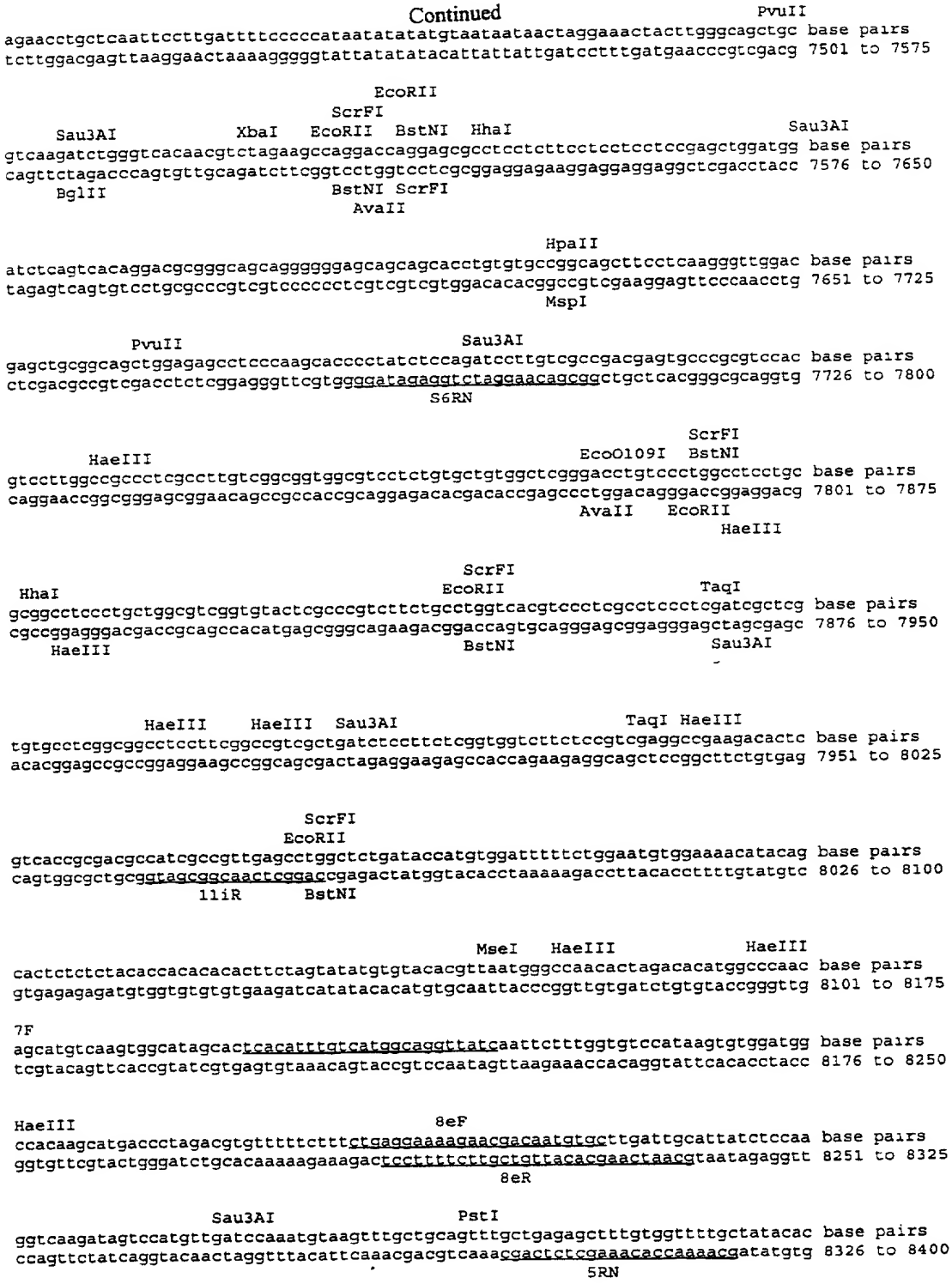
FIG. 5
Continued



14/39

FIG. 5

Continued



15/39

FIG. 5
Continued

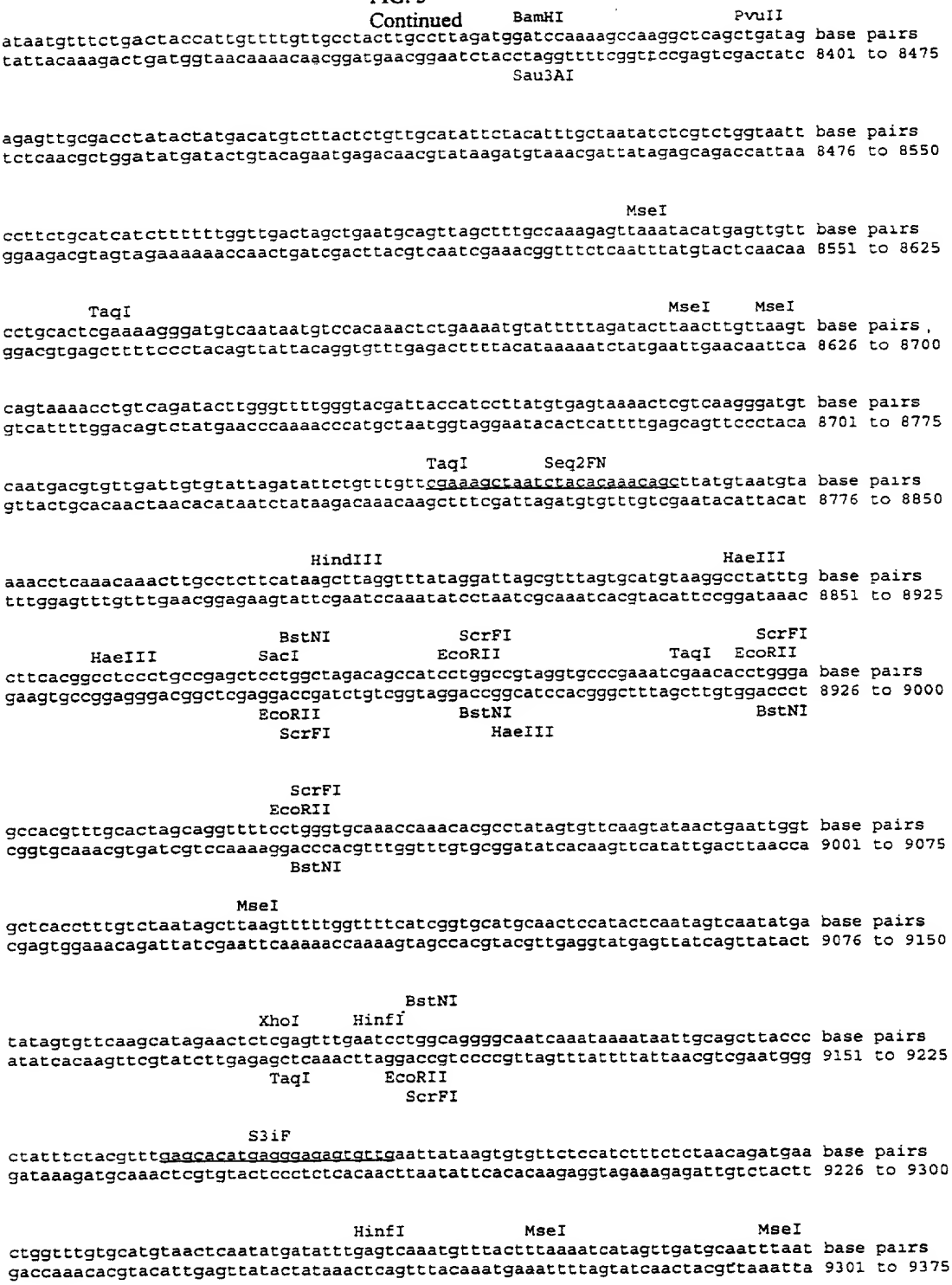


FIG. 5
Continued

```
aacatatttttttggtctcgtgtgagggagtgtacgtataactgaattgcacacatttccttatagcttaggttt base pairs
ttgtataaaaaaaccagagcacactccctcacatgcatattgacttaacgtgtgtaaaggaatatcgaatccaaa 9376 to 9450

                                                Sau3AI
ttgactgcaactgttggtgcatgtagctcaataactaaagttgatctggacagtctacagtgaataagtttgaca base pairs
aactgacgttgacaaccacgtacatcgagttattgatttcaactagacctgtcagatgtcacttattcaaactgt 9451 to 9525


cttgtaaaatgtgcatgtatttttacaaacgctggcacttttttcctaatagaaaatgggcagtcaggcagtgat base pairs
gaacattttacacgtacataaaaatgtttgcgaccgtgaaaaaaggattatcttttacccgtcagtccgtcacta 9526 to 9600

                                        Sau3AI
accgcttcgggtatttcttctgatgatgtggatctggagacgtcatctagtatgccaacgaggacagcaaccctt base pairs
tggcgaagcccataaagaagactactacacctagacctctgcagtagatcatacggttgctcctgtcgttgggaa 9601 to 9675

   Sau3AI
cttgatctgtattctggctgtgggggcatgtctactggtctttgcttgggtgcagctctttctggcttgaaactt base pairs
gaactagacataagaccgacacccccgtacagatgaccagaaacgaacccacgtcgagaaagaccgaactttgaa 9676 to 9750

                                                Sau3AI
gaaactgtaatcttctaactagtcatctgttggatagaatatgttcacgatctcagaacttattctattgttctg base pairs
ctttgacattagaagattgatcagtagacaacctatcttatacaagtgctagagtcttgaataagataacaagac 9751 to 9825

                                                MseI
gcttgcagcgatgggctgttgatttcaacagttttgcgtgccaaagtttaaaatataatcatccacagactgagg base pairs
cgaacgtcgctacccgacaactaaagttgtcaaaacgcacggtttcaaattttatattagtaggtgtctgactcc 9826 to 9900

                                HinfI
tatggatagtaaacttcatcttggattccatctgttctgtcagctactcttacaaagtgtctggatttttggatg base pairs
atacctatcatttgaagtagaacctaaggtagacaagacagtcgatgagaatgtttcacagacctaaaaacctac 9901 to 9975

                                                MseI
taggtgcgaaatgagaaagccgatgagtttcttgccctccttaaggaatgggcagttctatgcaaaaaatatgtc base pairs
atccacgctttactctttcggctactcaaagaacgggaggaattccttacccgtcaagatacgtttttttatacag 9976 to 10050

          HinfI                 Sau3AI
caagatgtggattcaaatttagcaagctcagaggatcaagcggatgaagacagccctcttgacaaggacgaattt base pairs
gttctacacctaagtttaaatcgttcgagtctcctagttcgcctacttctgtcgggagaactgttcctgcttaaa 10051 to 10125
                      Seq2RN

          HindIII                                                   MseI
gttgtagagaagcttgtcgggatatgttatggtggcagtgacagggaaaatggcatctattttaaggtacttcag base pairs
caacatctcttcgaacagccctatacaataccaccgtcactgtccctttaccgtagataaaattccatgaagtc 10126 to 10200

                          HinfI                 MseI
tgtcatttgttcatttctacttgattccaacaaaaaaatcaattacttaagcctgtcaaacgatggatatttctg base pairs
acagtaaacaagtaaagatgaactaaggttgtttttttagttaatgaattcggacagtttgctacctataaagac 10201 to 10275

                                  PstI                  HaeIII
tatattttgctgtaacgctagatttctgcaggtccagtgggaaggatacggccctgaggaggatacatgggaacc base pairs
atataaaacgacattgcgatctaaagacgtccaggtcacccttcctatgccgggactcctcctatgtacccttgg 10276 to 10350
                                  AvaII

gattgataacttgaggttagtgtatggtatatcgtctgcttgttgccttgtatacctatttgcatctaatccttg base pairs
ctaactattgaactccaatcacataccatatagcagacgaacaacgaacatatggataaacgtagattaggaac 10351 to 10425
                                  Seq1RN

                                                        HinfI
```

FIG. 5
Continued

ttttgcaaaccagtgactgcccgcagaaaattagagaatttgtacaagaagggcacaaaagaaagattctcccac base pairs
aaaacgtttggtcactgacgggcgtcttttaatctcttaaacatgttcttcccgtgttttctttctaagagggtg 10426 to 10500
S3eR

EcoRV
tgcctgtgagtatttagttcgttgtgattttgctcgctatttgtttagctcccctttttttatttggtgatatctg base pairs
acggacactcataaatcaagcaacactaaaacgagcgataaacaaatcgaggggaaaaaataaaccactatagac 10501 to 10575

HaeIII MseI
cctattttattctttcaaagggtgatgttgatgtcatttgtggaggcccaccatgccaaggtatcagtgggttta base pairs
ggataaaataagaaagtttcccactacaactacagtaaacacctccgggtggtacggttccatagtcacccaaat 10576 to 10650

3F
atcggtacagaaaccgtgatgagccactcaaagatgagaaaaacaaacaaatggtgactttcatggatattgtgg base pairs
tagccatgtctttggcactactcggtgagtttctactcttttttgtttgtttaccactgaaagtacctataacacc 10651 to 10725
3R

cgtacttgaagcccaagtatgttctcatggaaaatgtggtggacatactcaaatttgcggatggttacctaggaa base pairs
gcatgaacttcgggttcatacaagagtacctttttacaccacctgtatgagtttaaacgcctaccaatggatcctt 10726 to 10800

HhaI 1F
aatatgctttgagctgccttgttgctatgaagtaccaagcgcggcttggaatgatggtggctggttgctatggtc base pairs
ttatacgaaactcgacggaacaacgatacttcatggttcgcgccgaaccttactaccaccgaccaacgataccag 10801 to 10875

Race2A
tgccacagttcaggatgcgtgtgtacctctggggtgctctttctteeatggtctgttctgtaccttgctgtttta base pairs
acggtgtcaagtcctacgcacacatggagacccccacgagaaagaaggtaccagacaagacatggaacgacaaaat 10876 to 10950
Race2B

ScrFI
HinfI EcoRII
tatgcttcgctagattcatattgcactgttggctgctggctaaccaggtgtacgtgtatttgacaatttaggtgc base pairs
atacgaagcgatctaagtataacgtgacaaccgacgaccgattggtccacatgcacataaactgttaaatccaeg 10951 to 11025
BstNI

Race1A 4F
tccctaagtatcctctgcccaactatgatgttgtagtacgtggaggagcccctaatgcctttcggtgagtgcaat base pairs
agggattcatcggagacgggttgatactacaacatcatgcacctcctcggggattacggaaagccactcacgtta 11026 to 11100
Race1B RaceRT

cacaaaccactactatgaaatcatgtggaatgtgtaaaatacgctgaccaactgaatttgttgcagcaatgtatg base pairs
gtgtttggtgatgatactttagtacaccttacacattttatgcgactggttgacttaaacaacgtcgttacatac 11101 to 11175

gttgcatatgacgagacacaaaaaccatccctgaaaaaagccttgcttcttggcgatgcaatttcagatttacca base pairs
caacgtatactgctctgtgtttttggtagggacttttttcggaacgaagaaccgctacgttaaagtctaaatggt 11176 to 11250

MseI PstI
aaggcaagtgttctgtcaagttcatgcatttctcagtgagcatgctatttaactcttctctgcaggttcaaaatc base pairs
ttccgttcacaagacagttcaagtacgtaaagagtcactcgtacgataaattgagaagagacgtccaagttttag 11251 to 11325

EcoRI HhaI TaqI
accagcctaacgatgtgatggagtatggtggttccccaagaccgaattccagcgctacattcgactcagtcgtaa base pairs
tggtcggattgctacactacctcataccaccaaggggttctggcttaaggtcgcgatgtaagctgagtcagcatt 11326 to 11400
HinfI

HaeIII MseI
aggtaaaaaaccccgtgaactactactggttggccttcactacgaatatgttaggatttaatttcagaagaaccg base pairs
tccatttttttggggcacttgatgatgaccaaccggaagtgatgcttatacaatcctaaattaaagtcttcttggc 11401 to 11475

PstI AvaII
ccttttttttcttggtgcttcggtactactgcagcaagctcactcttattatcatgtcagacatgttggattggt base pairs
ggaaaaaaaagaaccacgaagccatgatgacgtcgttcgagtgagaataatagtacagtctgtacaacctaacca 11476 to 11550

FIG. 5
Continued

```
                    AvaII                Sau3AI                MseI
ccttcggtgaaggggggctggtccagatgaaggcaagctcttggatcaccagcctttacggcttaacaacgatgat base pairs
ggaagccacttcccccgaccaggtctacttccgttcgagaacctagtggtcggaaatgccgaattgttgctacta 11551 to 11625

                  HinfI
tatgagcgggttcaacagattcctgtcaagaaggttggtggcttggtcgcatttgtgccttcctttgttgttttt base pairs
atactcgcccaagttgtctaaggacagttcttccaaccaccgaaccagcgtaaacacggaaggaaacaacaaaaa 11626 to 11700

                                                                        5F
tcccctttctgaaacaatcatctctctttcctatgacagggagccaacttccgcgacctaaagggcgtgagggttg base pairs
aggggaagactttgttagtagagagaaaggatactgtccctcggttgaaggcgctggatttcccgcactcccaac 11701 to 11775

                        BamHI      TaqI
gagcaaacaatattgttgagtgggatccagaaatcgagcgtgtgaaactttcatctgggaaaccactggtatgtg base pairs
ctcgtttgttataacaactcaccctaggtctttagctcgcacactttgaaagtagacccttggtgaccatacac 11776 to 11850
                        Sau3AI

tgctatttccgtgctgttgtttcctataactgtgcaacatttactttcccatattcaaactcataactgacgaga base pairs
acgataaaggcacgacaacaaaggatattgacacgttgtaaatgaaagggtataagtttgagtattgactgctct 11851 to 11925

                  HinfI
tgctgcaactactgtaagattcatggctaacccatgacaacattttgcacacatctttgttatctaggttcctga base pairs
acgacgttgatgacattctaagtaccgattgggtactgttgtaaaacgtgtgtagaaacaatagatccaaggact 11926 to 12000

ctatgcaatgtcattcatcaagggcaaatcactcaagtaagtttcaaaacatttttttttgttttttgggggaaaa base pairs
gatacgttacagtaagtagttcccgtttagtgagttcattcaaagttttgtaaaaaaaacaaaaaaccccctttt 12001 to 12075

                                    HaeIII     HhaI
gtaggttattgtttacttgtgcttacatatgatgttgcaggccgtttgggcgcctgtggtgggacaagacagttc base pairs
catccaataacaaatgaacacgaatgtatactacaacgtccggcaaacccgcggacaccaccctgttctgtcaag 12076 to 12150

                              ScrFI
                            EcoRII              HaeIII
ctacagttgtaaccagagcagagcctcacaaccaggtcagcttcagaaaggccactccttttcgccaatccctgc base pairs
gatgtcaacattggtctcgtctcggagtgttggtccagtcgaagtctttccggtgaggaaaagcggttagggacg 12151 to 12225
                              BstNI

                              Sau3AI
atctgtatttactattagcgtgtgttcccatatgatcattaccgaacatgttgtccacacaggttataattcatc base pairs
tagacataaatgataatcgcacacaagggtatactagtaatggcttgtacaacaggtgtgtccaatattaagtag 12226 to 12300

ScrFI
  HinfI       EcoO109I    HpaII
cgactcaagcaagggtcctcactatccgggagaacgcaaggttacagggcttccccgattattaccgattgtttg base pairs
gctgagttcgttcccaggagtgataggccctcttgcgttccaatgtcccgaaggggctaataatggctaacaaac 12301 to 12375
             AvaII       MspI

 HaeIII                                   Sau3AI
gcccgatcaaggagaagtaagttcctgttttcaagttgcctgtaccagatctagtcactattgaaagttttcagc base pairs
cgggctagttcctcttcattcaaggacaaaagttcaacggacatggtctagatcagtgataactttcaaaagtcg 12376 to 12450
    Sau3AI                                BglII

agcaagccattcatcagttagttacagctcttgaaagccttacctctgaacatgtgtgctttctctgatggtgat base pairs
tcgttcggtaagtagtcaatcaatgtcgagaactttcggaatggagacttgtacacacgaaagagactaccacta 12451 to 12525

                                       MspI
                                       HpaII                          EcoRII
aggtacattcaagtcgggaacgcagtggctgtccctgttgcccgggcactgggctactgtctggggcaagcctac base pairs
tccatgtaagttcagcccttgcgtcaccgacagggacaacgggcccgtgacccgatgacagaccccgttcggatg 12526 to 12600
                                       ScrFI
                                        SmaI
```

FIG. 5
Continued

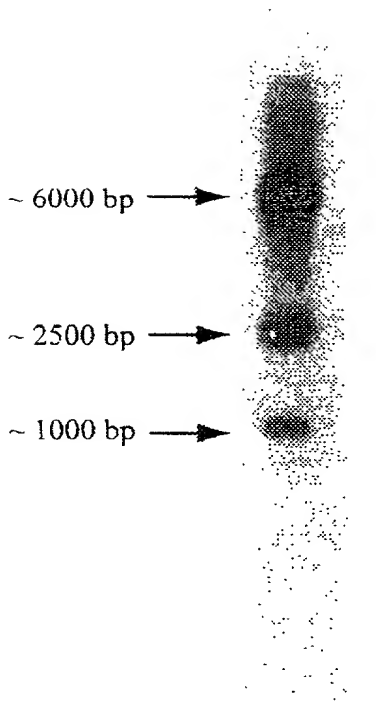
```
ScrFI
      HinfI                                PvuII
ctgggtgaatctgaggggagtgaccctctgtaccagctgcctccaagtttcacctctgttggaggacgcactgcg base pairs
gacccacttagactcccctcactgggagacatggtcgacggaggttcaaagtggagacaacctcctgcgtgacgc 12601 to 12675
 BstNI

          EcoO109I                    PstI                                  Sau3AI
gggcaggcgagggcctcttcctgttggcacccctgcaggggaggtagttgagcagtaaaaggatgacagatctga base pairs
cccgtccgctcccggagaaggacaaccgtggggacgtcccctccatcaactcgtcatttcctactgtctagact 12676 to 12750
             HaeIII                                                 1R    BglII
                                          TaqI
gctgagctgggcaacatccagcggcaggagcatttctggttcggttcgattcgggctcacga  base pairs
cgactcgacccgttgtaggtcgccgtcctcgtaaagaccaagccaagctaagcccgagtgct  12751 to 12812
                                        HinfI
```

FIG. 6

PROCESS	WORLD WIDE WEB SITE
sequence format conversion	http://dot.imgen.bcm.tmc.edu:9331/seq-util/Options/readseq.html
reverse complementation	http://dot.imgen.bcm.tmc.edu:9331/seq-util/Options/revcomp.html
sequence translation	http://dot.imgen.bcm.tmc.edu:9331/seq-util/Options/sixframe.html
protein information	http://www.expasy.ch/tools
sequence alignments using Clustal W	http://dot.imgen.bcm.tmc.edu:9331/multi-align/Options/clustalw.html
sequence comparisons using BLAST 2.0	http://www.ncbi.nlm.nih.gov/gorf/bl2.html
sequence searches using BLAST 2.0	http://www.ncbi.nlm.nih.gov/blast/blast.cgi?Jform=0

FIG. 7



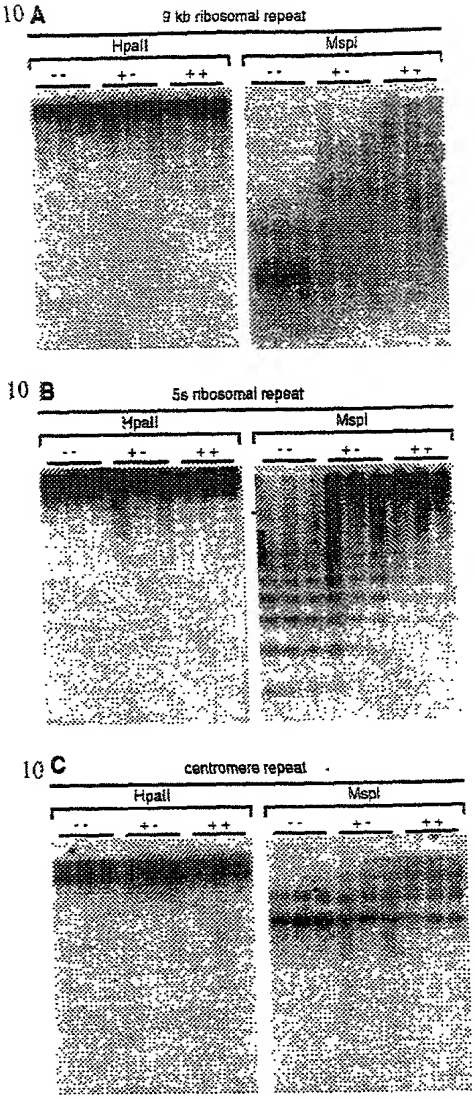
22/39

FIG. 8

FIG. 9

	SAM binding		Cytosine binding	
Motif	M.*Hhal*	*zmet2a*	M.*Hhal*	*zmet2a*
I	Phe18	Try347		
II	Glu40	Gln407		
	Trp41	Trp408		
III	Asp60	Asp428		
IV	Pro80	Pro516	Phe79	Pro515
	Gln82	Gln82	Cys81	Cys517
V	Leu100	Val542		
VI			Glu119	Glu559
			Asn120	Asn560
			Val121	Val561
VIII			Arg165	Arg605
X	Asn304	Asn851		

FIG. 10



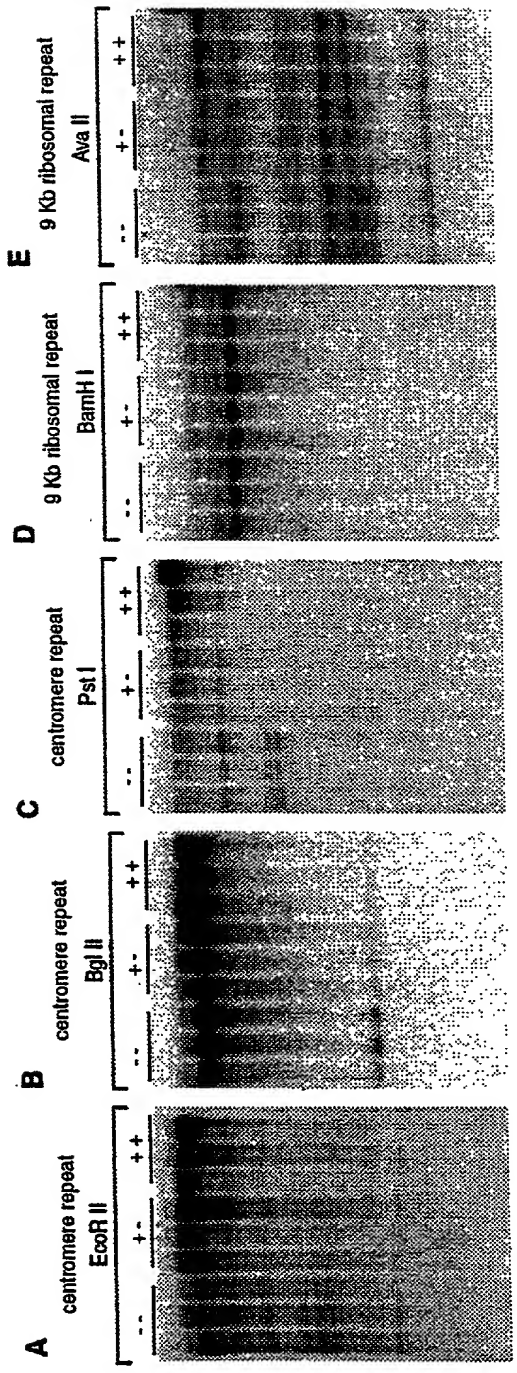

A
centromere repeat
EcoR II
B
centromere repeat
Bgl II
C
centromere repeat
Pst I
D
9 Kb ribosomal repeat
BamH I
E
9 Kb ribosomal repeat
Ava II
- -
+ -
+ +

FIG. 11

26/39

FIG. 12

GENOTYPE	NUMBER OF PLANTS	TOTAL 5mCytosine (%)	% WT levels	% decrease
wild type	3	34.40 ± 0.55	100	0.0
heterozygous zmet2a-mu1	7	32.00 ± 0.90	93.0	7.0
homozygous zmet2a-mu1	5	30.40 ± 0.19	88.4	11.6

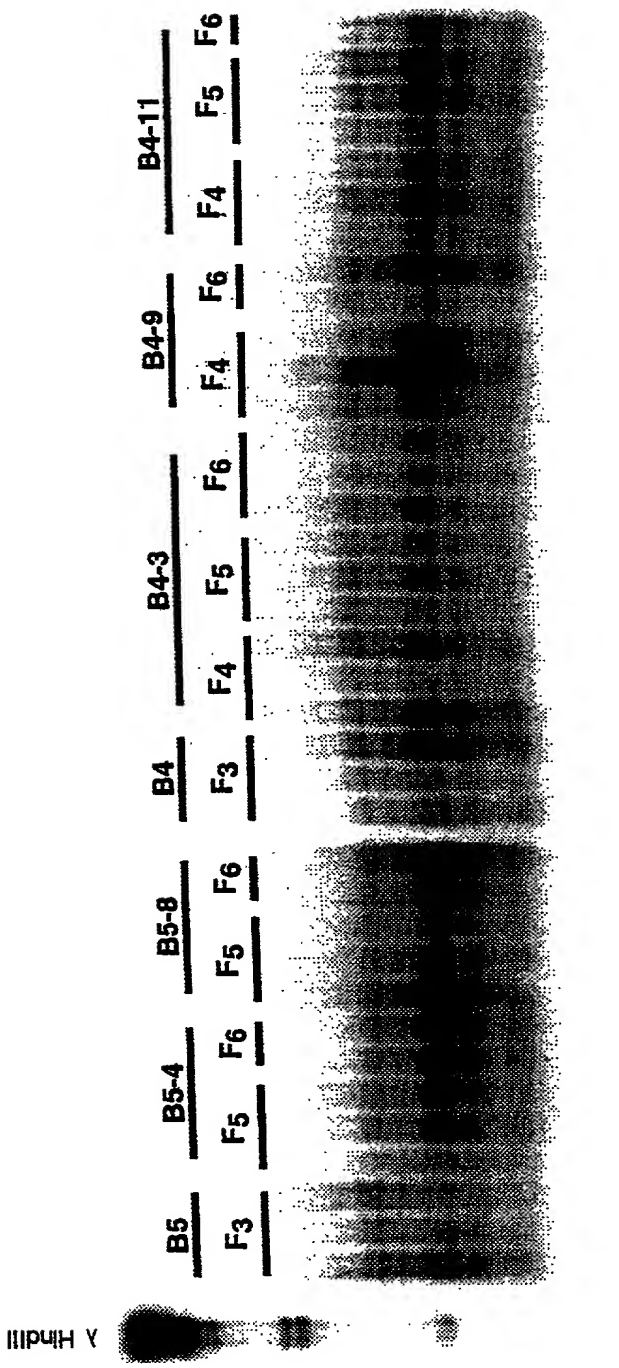

λ HindIII
B5
F3
B5-4
F5
F6
B5-8
F5
F6
B4
F3
B4-3
F4
F5
F6
B4-9
F4
F6
B4-11
F4
F5
F6

FIG. 13

28/39

FIG. 14

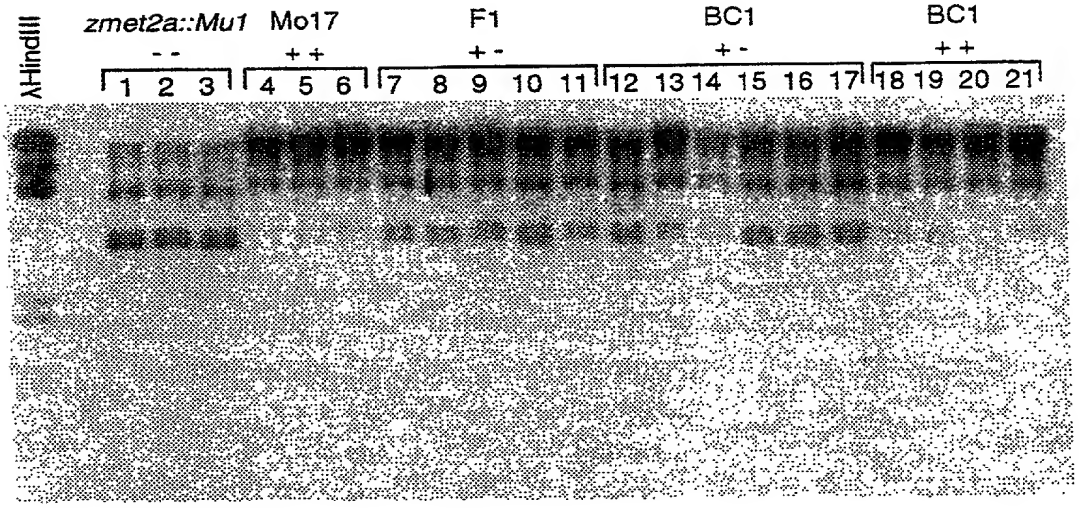

λ-HindIII
zmet2a::Mu1
Mo17
F1
BC1
BC1
- -
+ +
+ -
+ -
+ +
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

FIG. 15

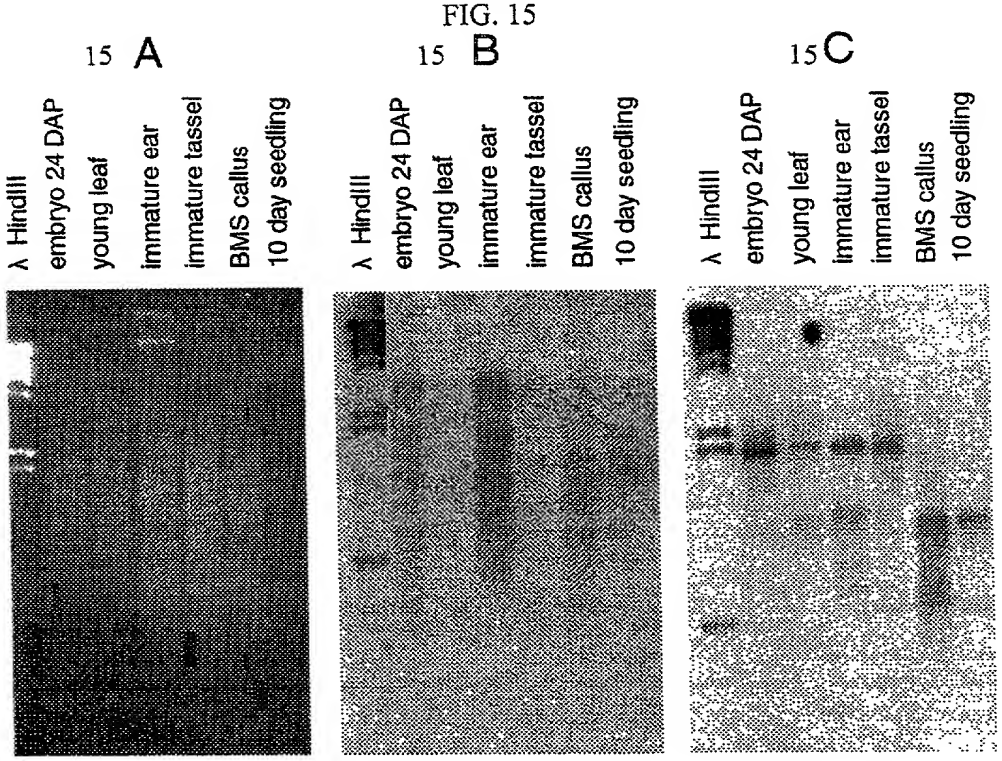

15 A
λ HindIII
embryo 24 DAP
young leaf
immature ear
immature tassel
BMS callus
10 day seedling
15 B
λ HindIII
embryo 24 DAP
young leaf
immature ear
immature tassel
BMS callus
10 day seedling
15 C
λ HindIII
embryo 24 DAP
young leaf
immature ear
immature tassel
BMS callus
10 day seedling

FIG.16

5' LTR

catgc**TGT**TGGGCCATGTGTCTAGTGTTGGCCCATTAACGTGTACA
CATATACTAGAAGTGTGTGTGGTGTAGAGAGAGTGCTGTATGTTTT
CCACATTCCAGAAAAATCC**ACA**TGGTATCAGAGCCAGG
PBS

3' LTR

PPT
GAGGGGGAG**TGT**TGGGCCATGTGTCTAGTGTTGGCCCATTAACGTG
TACACATATACTAGAAGTGTGTGTGGTGTAGAGAGAGTGCTGTATG
TTTTCCACATTCCAGAAAAATCC**ACA**catgc

FIG. 17

Gag **Protease**

```
SPRITE-1   - CYNCGNVGHIARNC    TQVTQLKWILDSGASKH
hopscotch  - CQVCSRVGHTALNC    QNGSNVPWYTDTGATDH
retrofit   - CQVCFKRGHTAADC    SYGIDTNWYIDTGATDH
arabpolprt- CSNCGRTGHEKKEC     GKTKLGDIILDSGASHH
copia      - CHHCGREGHIKKDC     SVMDNCGFVLDSGASDH
```

Integrase

```
SPRITE-1   - QVKILRPDN-GTEYVNKGFNAFLSRNGILHQTSCPDTPPQNGVAERKNRHILE
hopscotch  - KIIAFQSDW-GGE--YEKLNAHFKTIGIHHQVSCPHTHQQNGAAERKHRHIVE
retrofit   - KIIAMQTDWRGGR--YQKLNSFFAQIGLIIMCHVLTLIRQNGSAERKHRHIVE
arabpolprt- TVKMVRSDN-GTE--FMCLSSYFRENGIIHQTSCVGTPQQNGRVERKHRHILN
copia      - KVVYLYIDN-GREYLSNEMRQFCVKKGISYHLTVPHTPQLNGVSERMIRTITE
```

Reverse Transcriptase

```
SPRITE-1   - RYKARLVARGYSQTYGIDYDETFAPVAKMSTVRTLISCAANFGWPLYQLDVKNAFLHGDLQEEVYMEIPPG(59)AILAVYVDDIII
hopscotch  - RLKARLVAKGFKQQYGIDYDDTFSPVVKHSTIRLVLSLAVSQKWSLRQLDVQNAFLHGILEETVYMKQPPG(59)IYILVYVDDIII
Retrofit   - RYKARLVAKGFKQRYGIDYEDTFSPVVKAATIRIILSIAVSRGWSLRQLDVQNAFLHGFLEEEVYMQQPPG(59)MFVLVYVDDIIV
Arabpolprt- RYKARLVVQGNKQVEGEDYKETFAPVVRMTTVRTLLRNVAANQWEVYQMDVHNAFLHGDLEEEVYMKLPPG(59)LRVLIYVDDLLI
copia      - RYKARLVARGFTQKYQIDYEETFAPVARISSFRFILSLVIQYNLKVHQMDVKTAFLNGTLKEEIYMRLPQG(59)IYVLLYVDDVVI
```

RNase H

```
SPRITE-1   - DADWGSCLDDRRSTSGYCVFVGG-NLVSWRSKKQSVVSRSTAEAEYRAMALAICEMLWIKGLL(25)NPVQHDRTKHVEIDRFF
hopscotch  - DADWAGCPDDRKSTGGYALFLGP-NLISWNSKKQSTVSRSSTEAEYKAMANATAEVIWLQSLL(25)KPIFNARTKHIEVDFHF
retrofit   - DADWAGSIDDRKSTGGFAVFLGS-NLVSWSARKQPTVSRSSTEAEYKAVANTTAELIWVQTLL(25)NPVFHARTKHIEVDYHF
arabpolprt- DSDWQSCPLTRRSISAYVVLLGG-SPISWKTKKQDTVSHSSAEAEYRAMSYALKEIKWLRKLL(25)NPVFHERTKHIESDCHS
copia      - DSDWAGSEIDRKSTTGYLFKMFDFNLICWNTKRQNSVAASSTEAEYMALFEACREALWLKFLL(25)NPSCHKRAKHIDIKYHF
```

FIG. 18

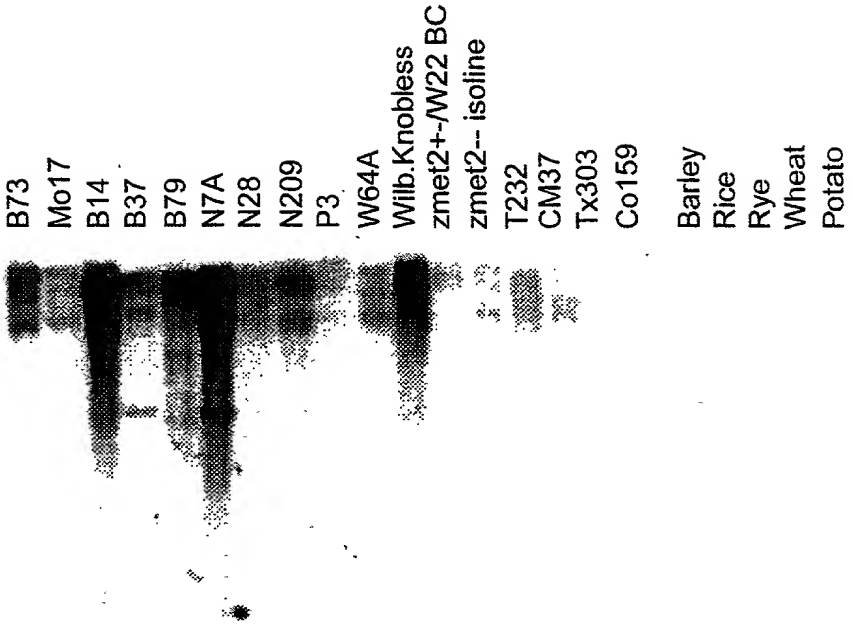

B73
Mo17
B14
B37
B79
N7A
N28
N209
P3
W64A
Wilb.Knobless
zmet2+-/W22 BC
zmet2-- isoline
T232
CM37
Tx303
Co159
Barley
Rice
Rye
Wheat
Potato

FIG. 19

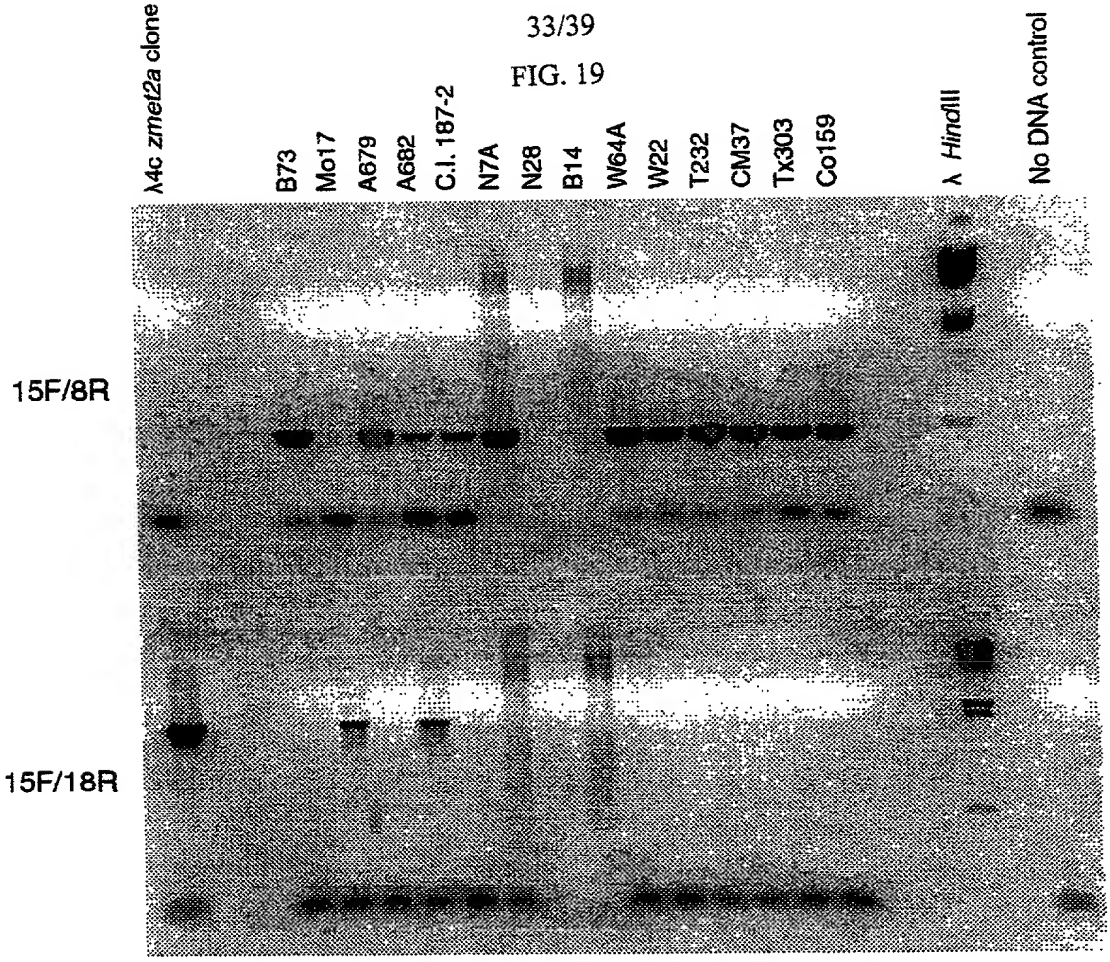

λ4c zmet2a clone
B73
Mo17
A679
A682
C.I. 187-2
N7A
N28
B14
W64A
W22
T232
CM37
Tx303
Co159
λ HindIII
No DNA control
15F/8R
15F/18R

FIG. 20

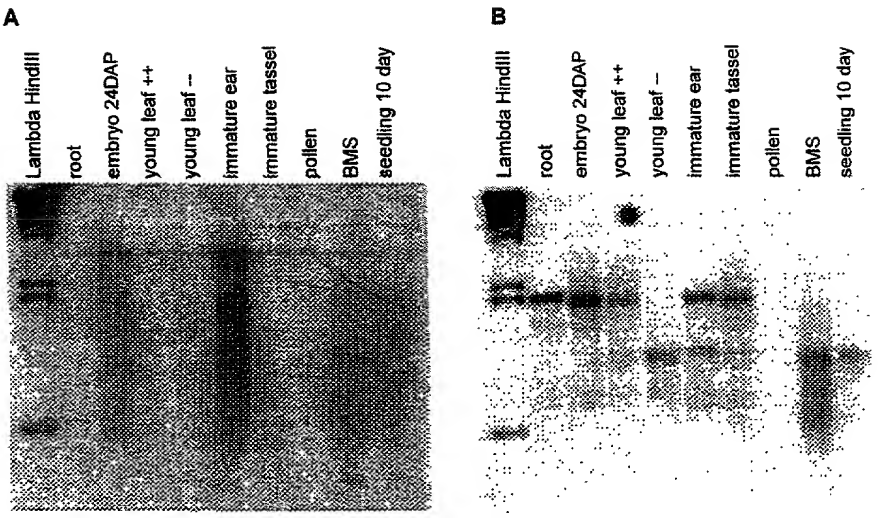
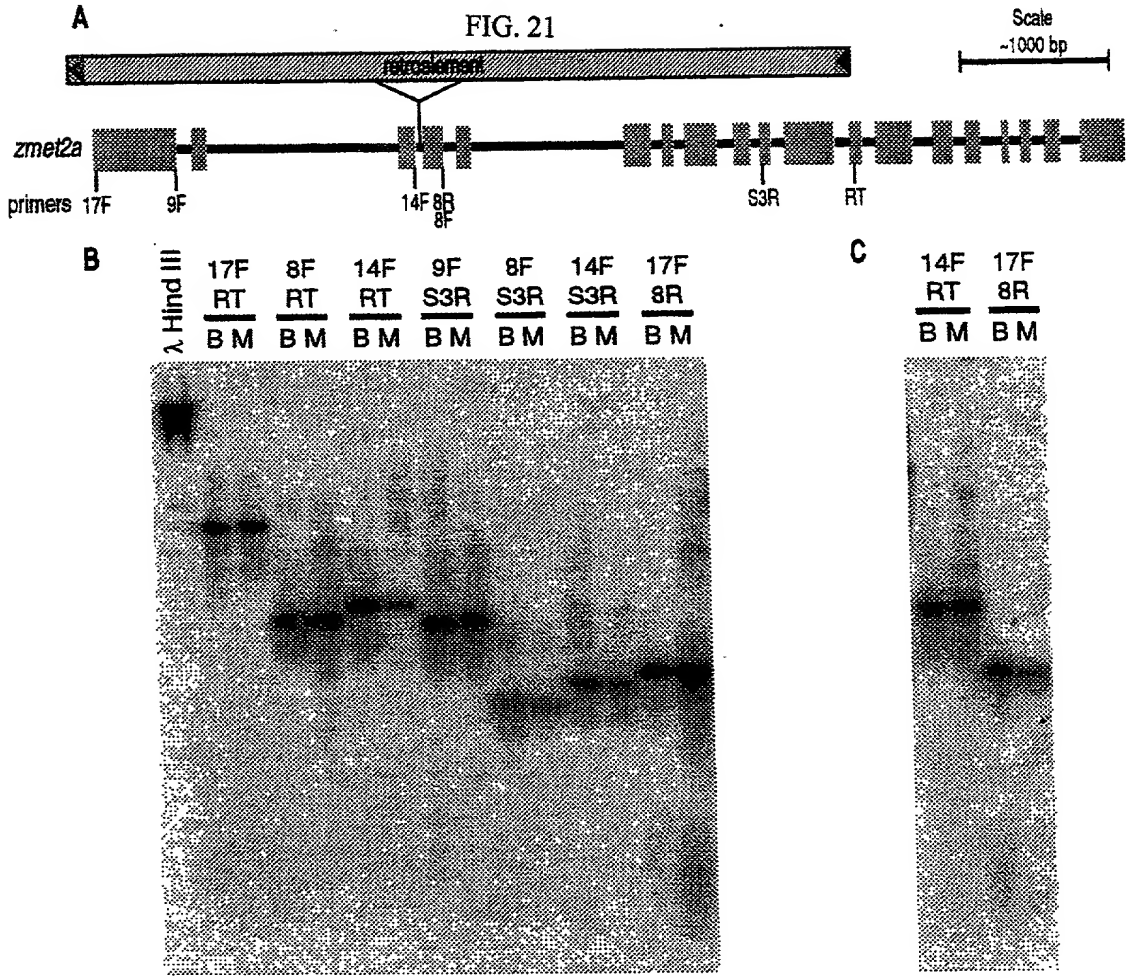

A
Lambda HindIII
root
embryo 24DAP
young leaf ++
young leaf --
immature ear
immature tassel
pollen
BMS
seedling 10 day
B
Lambda HindIII
root
embryo 24DAP
young leaf ++
young leaf --
immature ear
immature tassel
pollen
BMS
seedling 10 day

FIG. 21


A
retrotransposon
Scale
~1000 bp
zmet2a
primers 17F
9F
14F
8R
8F
S3R
RT
B
λ Hind III
17F RT
8F RT
14F RT
9F S3R
8F S3R
14F S3R
17F 8R
B M
C
14F RT
17F 8R

36/39

FIG. 22

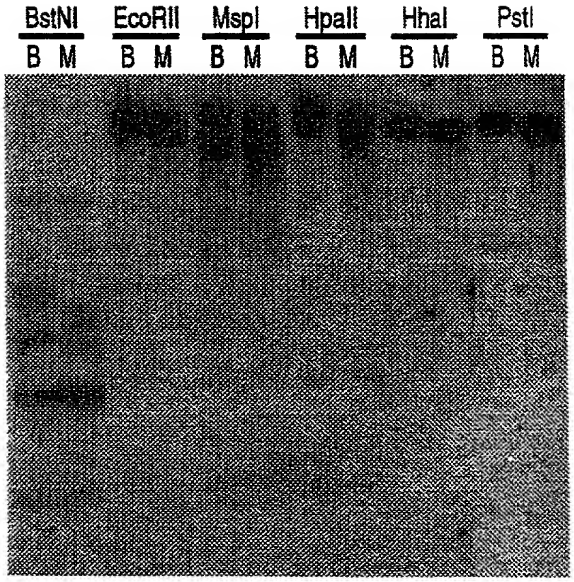


FIG. 23

GGGAATTCGATTACTCACTATAGCGCTCGAGCGGCCGCCCGGGCAGGTTCGAAAACCATC
AACCTAACGATGTAATGGAGTATGGTGGTTCCCCCAAGACAGAGTTCCAGCGCTACATTC
GACTTGGTCGTAAAGACATGTTGGATTGGTCGTTTGGTGAGGAGGCTGGTCCAGATGAAG
GCAAGCTCTTGGATCACCAGCCCTTACGGCTTAACAATGATGATTATGAGCGGGTTAAGC
AAATTCCTGTCAAGAAGGGAGCCAACTTCCGTGACCTAAAGGGTGTCAAGGTTGGAGCAA
ATAATGTTGTTGAGTGGGATCCAGAAGTCGAACGTGTGTACCTTTCGTCTGGGAAACCAC
TGGTTCCTGACTATGCGATGTCATTCATCAAGGGCAAATCACTCAAGCCATTCGGGCGCC
TGTGGTGGGACCAGACGGTTcCTACAGTTGTGACCAGAGCAGAGCCTCATAACCAGGTTAT
ATTGCATCCGACTCAAGCAAGAGTCTTGACTATCCGGGAGAACGCAAGGTTACAGGGCTT
CCCCGATTACTACCGATTGTTTGGACCGATCAAGGAGAAGTATATTCAAGTCGGGAACGC
AGTGGCAGTCCCTGTTGCACGGGCACTGGGCTACTGTCTGGGTCAAGCCTACCTGGGTGA
ATCTGACGGGAGTCAGCCTCTGTACCAGCTGCCTGCAAGTTTTACCTCTGTGGGGCGAAC
CGCGGTTCAGGCGAATGCCGCTTCTGTTGGCACTCCTGCGGGGGAGGTAGTCGAGCAGTA
AAAGGATAGCGGAGCAACCCTGGTTGGTATTTTGATTCGAGCCCATCCAGTAGCATGTTT
ACCAATAAATAATCATTGGTCGTGCTGATTCTTATGGTTGGAGATGAATGTATGTAGGGT
GTACTCGAGCTCGAGTGCTTGTTGTACTGTAGGTTGAGGTTTCTCATCCATTGGCCTGCC
TATTTGTGGATGACGTTTCATTTCAGATTAGCAATGTGCTTATTTAAGGTTTCGTCATGT
ACCTGTATTCTACAATCCACTATTGTTTCCAAAGACAGCATTTGATCCTTAAAAAAAACT
GTAAAAAAAAAAAAAACAGTGCCCGAAAAGCCGCAAAAAAAAAAAAAAAAAAAACCTGCCC
GGGCGGCCGCTCGAGCCCTATAGTGAGTAATCGAATTCCC

FIG. 24

EFDYSL*RSSGRPGRFENHQPNDVMEYGGSPKTEFQRYIRLGRKDMLDWS
FGEEAGPDEGKLLDHQPLRLNNDDYERVKQIPVKKGANFRDLKGVKVGAN
NVVEWDPEVERVYLSSGKPLVPDYAMSFIKGKSLKPFGRLWWDQTVPTVV
TRAEPHNQVILHPTQARVLTIRENARLQGFPDYYRLFGPIKEKYIQVGNA
VAVPVARALGYCLGQAYLGESDGSQPLYQLPASFTSVGRTAVQANAASVG
TPAGEVVEQ*

39/39

FIG. 25

```
667  KVQNHQPNDVMEYGGSPKTEFQRYIRLSRKDMLDWSFGEGAGPDEGKLLDHQPLRLNNDD  726
     +  +NHQPNDVMEYGGSPKTEFQRYIRL RKDMLDWSFGE AGPDEGKLLDHQPLRLNNDD
15   RFENHQPNDVMEYGGSPKTEFQRYIRLGRKDMLDWSFGEEAGPDEGKLLDHQPLRLNNDD  74

727  YERVQQIPVKKGANFRDLKGVRVGANNIVEWDPEIERVKLSSGKPLVPDYAMSFIKGKSL  786
     YERV+QIPVKKGANFRDLKGV+VGANN+VEWDPE+ERV LSSGKPLVPDYAMSFIKGKSL
75   YERVKQIPVKKGANFRDLKGVKVGANNVVEWDPEVERVYLSSGKPLVPDYAMSFIKGKSL  134

787  KPFGRLWWDETVPTVVTRAEPHNQVIIHPTQARVLTIRENARLQGFPDYYRLFGPIKEKY  846
     KPFGRLWWD+TVPTVVTRAEPHNQVI+HPTQARVLTIRENARLQGFPDYYRLFGPIKEKY
135  KPFGRLWWDQTVPTVVTRAEPHNQVILHPTQARVLTIRENARLQGFPDYYRLFGPIKEKY  194

847  IQVGNAVAVPVARALGYCLGQAYLGESEGSDPLYQLPPSFTSVGGRTAGQARASPVGTPA  906
     IQVGNAVAVPVARALGYCLGQAYLGES+GS PLYQLP SFTSV GRTA QA A+ VGTPA
195  IQVGNAVAVPVARALGYCLGQAYLGESDGSQPLYQLPASFTSV-GRTAVQANAASVGTPA  253

907  GEVVEQ 912
     GEVVEQ
254  GEVVEQ 259
```

DECLARATION AND POWER OF ATTORNEY FOR UTILITY OR DESIGN PATENT APPLICATION (37 CFR 1.63)	Attorney Docket No.: WIS4987P0052US
	First Named Inventor: Shawn M. Kaeppler *et al.*
	COMPLETE IF KNOWN
☐ Declaration Submitted With Initial Filing ☒ Declaration Submitted After Initial Filing (surcharge (37 CFR 1.16(a)) required	Application Number: 09/914,001
	Filing Date: August 20, 2001
	Group Art Unit:
	Examiner Name:

As a below-named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name.

I believe I am the original, first, and sole inventor (if only one name is listed) or an original, first and joint inventor (if plural names are listed) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **Class II DNA Methyltransferases of Zea mays**, the specification of which:

☐ is attached hereto; or

☒ was filed on March 10, 2000 as Application Serial No. PCT/US00/06456 and was amended on ____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose all information which is material to patentability as defined in 37 CFR. 1.56, including for continuation-in-part applications, material information which became available between the filing date of the prior application and the national or PCT international filing date of the continuation-in-part application.

I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or 365(b) of any foreign application(s) for patent or inventor's certificate, or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Numbers	Country	Foreign Filing Date (MM/DD/YYY)	Priority Not Claimed	Certified Copy Attached?	
				YES	NO
			☐	☐	☐
			☐	☐	☐
			☐	☐	☐

☐ Additional foreign application numbers are listed on a supplemental priority data sheet attached hereto.

I hereby claim the benefit of any United States application(s) listed below.

Application Number(s)	Filing Date	
60/123,888 60/169,858	11 March 1999 09 December 1999	☐ Additional application numbers are listed on a supplemental priority data sheet attached hereto.

The undersigned hereby authorizes the U.S. attorney(s) or agent(s) named herein to accept and follow instructions from the assignee, if any, of the undersigned or from as to any action to be taken in the Patent and Trademark Office regarding this application without direct communication between the U.S. attorney(s) or agent(s) and the undersigned. In the event of a change in the persons from whom instructions may be taken, the U.S. attorney(s) or agent(s) named herein will be so notified by the undersigned.

As a named inventor, I hereby appoint the following registered practitioner(s) to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

Lawrence J. Chapa	Reg. No. 39,135	Martin L. Katz	Reg. No. 25,011	Keith V. Rockey	Reg. No. 24,713
Randall T. Erickson	Reg. No. 33,872	Kathleen A. Lyons	Reg. No. 31,852	Thomas I. Ross	Reg. No. 29,275
Stephen D. Geimer	Reg. No. 28,846	John P. Milnamow	Reg. No. 20,635	Joel E. Siegel	Reg. No. 25,440
H. Vincent Harsha	Reg. No. 18,045	Lisa V. Mueller	Reg. No. 38,978	Paul M. Vargo	Reg. No. 29,116
Allen J. Hoover	Reg. No. 24,103	Paul M. Odell	Reg. No. 28,332		

whose mailing address for this application is: ROCKEY, MILNAMOW & KATZ, LTD.
Two Prudential Plaza - Suite 4700
180 North Stetson Avenue
Chicago, Illinois 60601
Telephone: (312) 616-5400
Facsimile: (312) 616-5460

Customer Number (01942) and/or Bar Code Label:

I hereby declare that all statements made herein of my knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Name of Sole or First Inventor:	Shawn M. Kaeppler	
Citizenship:	US	
Residence:	5290 County Highway A, Oregon, Wisconsin 53575	
Post Office Address *(if different)*:	same as above	
Signature: Shawn Kaeppler		Date: 9/20/01
☐ A petition has been filed for this unsigned inventor.		

Name of Additional Inventor, if any: 2nd	Nathan M. Springer
Citizenship:	US
Residence:	918 Washington Street, Northfield, MN 55057 MN
Post Office Address *(if different)*:	Same as above
Signature: Nathan Springer	Date: 9/17/01
☐ A petition has been filed for this unsigned inventor.	

3rd **Name of Additional Inventor**, if any:	Michael G. Muszynski
Citizenship:	US
Residence:	5505 Shriver Avenue #2, Johnston, IA 50131 IA
Post Office Address *(if different)*:	same as above
Signature:	Date:
☐ A petition has been filed for this unsigned inventor.	

4th **Name of Additional Inventor**, if any:	Charles M. Papa
Citizenship:	US
Residence:	903 Beacon Street #1, Madison, WI 53715 WI
Post Office Address *(if different)*:	
Signature:	Date:
☐ A petition has been filed for this unsigned inventor.	

Name of Additional Inventor, if any:	
Citizenship:	
Residence:	
Post Office Address *(if different)*:	
Signature:	Date:
☐ A petition has been filed for this unsigned inventor.	

P99/7205

DECLARATION AND POWER OF ATTORNEY FOR UTILITY OR DESIGN PATENT APPLICATION (37 CFR 1.63)	Attorney Docket No.: WIS4987P0052US
	First Named Inventor: Shawn M. Kaeppler *et al.*
	COMPLETE IF KNOWN
☐ Declaration Submitted With Initial Filing ☒ Declaration Submitted After Initial Filing (surcharge (37 CFR 1.16(a)) required	Application Number: 09/914,001
	Filing Date: August 20, 2001
	Group Art Unit:
	Examiner Name:

As a below-named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name.

I believe I am the original, first, and sole inventor (if only one name is listed) or an original, first and joint inventor (if plural names are listed) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **Class II DNA Methyltransferases of Zea mays**, the specification of which:

☐ is attached hereto; or

☒ was filed on March 10, 2000 as Application Serial No. PCT/US00/06456 and was amended on ____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose all information which is material to patentability as defined in 37 CFR. 1.56, including for continuation-in-part applications, material information which became available between the filing date of the prior application and the national or PCT international filing date of the continuation-in-part application.

I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or 365(b) of any foreign application(s) for patent or inventor's certificate, or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Numbers	Country	Foreign Filing Date (MM/DD/YYY)	Priority Not Claimed	Certified Copy Attached?	
				YES	NO
			☐	☐	☐
			☐	☐	☐
			☐	☐	☐

☐ Additional foreign application numbers are listed on a supplemental priority data sheet attached hereto.

I hereby claim the benefit of any United States application(s) listed below.

Application Number(s)	Filing Date	
60/123,888 60/169,858	11 March 1999 09 December 1999	☐ Additional application numbers are listed on a supplemental priority data sheet attached hereto.

The undersigned hereby authorizes the U.S. attorney(s) or agent(s) named herein to accept and follow instructions from the assignee, if any, of the undersigned or from as to any action to be taken in the Patent and Trademark Office regarding this application without direct communication between the U.S. attorney(s) or agent(s) and the undersigned. In the event of a change in the persons from whom instructions may be taken, the U.S. attorney(s) or agent(s) named herein will be so notified by the undersigned.

As a named inventor, I hereby appoint the following registered practitioner(s) to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

Lawrence J. Chapa	Reg. No. 39,135	Martin L. Katz	Reg. No. 25,011	Keith V. Rockey	Reg. No. 24,713
Randall T. Erickson	Reg. No. 33,872	Kathleen A. Lyons	Reg. No. 31,852	Thomas I. Ross	Reg. No. 29,275
Stephen D. Geimer	Reg. No. 28,846	John P. Milnamow	Reg. No. 20,635	Joel E. Siegel	Reg. No. 25,440
H. Vincent Harsha	Reg. No. 18,045	Lisa V. Mueller	Reg. No. 38,978	Paul M. Vargo	Reg. No. 29,116
Allen J. Hoover	Reg. No. 24,103	Paul M. Odell	Reg. No. 28,332		

whose mailing address for this application is: ROCKEY, MILNAMOW & KATZ, LTD.
Two Prudential Plaza - Suite 4700
180 North Stetson Avenue
Chicago, Illinois 60601
Telephone: (312) 616-5400
Facsimile: (312) 616-5460

Customer Number (01942) and/or Bar Code Label:

I hereby declare that all statements made herein of my knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Name of Sole or First Inventor:	Shawn M. Kaeppler	
Citizenship:	US	
Residence:	5290 County Highway A, Oregon, Wisconsin 53575	
Post Office Address *(if different):*	same as above	
Signature:		Date:
☐ A petition has been filed for this unsigned inventor.		

Name of Additional Inventor, if any:	Nathan M. Springer
Citizenship:	US
Residence:	918 Washington Street, Northfield, MN 55057
Post Office Address *(if different)*:	Same as above
Signature:	Date:
☐ A petition has been filed for this unsigned inventor.	

Name of Additional Inventor, if any:	Michael G. Muszynski
Citizenship:	US
Residence: 1575 NW 75th, CLIVE, IA 50325	~~5505 Shriver Avenue #2, Johnston, IA 50131~~
Post Office Address *(if different)*:	same as above
Signature: Michael G Muszynski	Date: 9/17/01
☐ A petition has been filed for this unsigned inventor.	

Name of Additional Inventor, if any:	Charles M. Papa
Citizenship:	US
Residence:	903 Beacon Street #1, Madison, WI 53715
Post Office Address *(if different)*:	
Signature:	Date:
☐ A petition has been filed for this unsigned inventor.	

Name of Additional Inventor, if any:	
Citizenship:	
Residence:	
Post Office Address *(if different)*:	
Signature:	Date:
☐ A petition has been filed for this unsigned inventor.	

DECLARATION AND POWER OF ATTORNEY FOR UTILITY OR DESIGN PATENT APPLICATION (37 CFR 1.63)	Attorney Docket No.: WIS4987P0052US
	First Named Inventor: Shawn M. Kaeppler *et al.*
	COMPLETE IF KNOWN
☐ Declaration Submitted With Initial Filing ☒ Declaration Submitted After Initial Filing (surcharge (37 CFR 1.16(a)) required	Application Number: 09/914,001
	Filing Date: August 20, 2001
	Group Art Unit:
	Examiner Name:

As a below-named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name.

I believe I am the original, first, and sole inventor (if only one name is listed) or an original, first and joint inventor (if plural names are listed) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **Class II DNA Methyltransferases of Zea mays**, the specification of which:

☐ is attached hereto; or

☒ was filed on March 10, 2000 as Application Serial No. PCT/US00/06456 and was amended on ____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose all information which is material to patentability as defined in 37 CFR. 1.56, including for continuation-in-part applications, material information which became available between the filing date of the prior application and the national or PCT international filing date of the continuation-in-part application.

I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or 365(b) of any foreign application(s) for patent or inventor's certificate, or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Numbers	Country	Foreign Filing Date (MM/DD/YYY)	Priority Not Claimed	Certified Copy Attached?	
				YES	NO
			☐	☐	☐
			☐	☐	☐
			☐	☐	☐

☐ Additional foreign application numbers are listed on a supplemental priority data sheet attached hereto.

I hereby claim the benefit of any United States application(s) listed below.

Application Number(s)	Filing Date	
60/123,888 60/169,858	11 March 1999 09 December 1999	☐ Additional application numbers are listed on a supplemental priority data sheet attached hereto.

The undersigned hereby authorizes the U.S. attorney(s) or agent(s) named herein to accept and follow instructions from the assignee, if any, of the undersigned or from as to any action to be taken in the Patent and Trademark Office regarding this application without direct communication between the U.S. attorney(s) or agent(s) and the undersigned. In the event of a change in the persons from whom instructions may be taken, the U.S. attorney(s) or agent(s) named herein will be so notified by the undersigned.

As a named inventor, I hereby appoint the following registered practitioner(s) to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

Lawrence J. Chapa	Reg. No. 39,135	Martin L. Katz	Reg. No. 25,011	Keith V. Rockey	Reg. No. 24,713
Randall T. Erickson	Reg. No. 33,872	Kathleen A. Lyons	Reg. No. 31,852	Thomas I. Ross	Reg. No. 29,275
Stephen D. Geimer	Reg. No. 28,846	John P. Milnamow	Reg. No. 20,635	Joel E. Siegel	Reg. No. 25,440
H. Vincent Harsha	Reg. No. 18,045	Lisa V. Mueller	Reg. No. 38,978	Paul M. Vargo	Reg. No. 29,116
Allen J. Hoover	Reg. No. 24,103	Paul M. Odell	Reg. No. 28,332		

whose mailing address for this application is: ROCKEY, MILNAMOW & KATZ, LTD.
Two Prudential Plaza - Suite 4700
180 North Stetson Avenue
Chicago, Illinois 60601
Telephone: (312) 616-5400
Facsimile: (312) 616-5460

Customer Number (01942) and/or Bar Code Label:

I hereby declare that all statements made herein of my knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Name of Sole or First Inventor:	Shawn M. Kaeppler	
Citizenship:	US	
Residence:	5290 County Highway A, Oregon, Wisconsin 53575	
Post Office Address *(if different):*	same as above	
Signature:		Date:
☐ A petition has been filed for this unsigned inventor.		

Name of Additional Inventor, if any:	Nathan M. Springer
Citizenship:	US
Residence:	918 Washington Street, Northfield, MN 55057
Post Office Address *(if different)*:	Same as above
Signature:	Date:
☐ A petition has been filed for this unsigned inventor.	

Name of Additional Inventor, if any:	Michael G. Muszynski
Citizenship:	US
Residence:	5505 Shriver Avenue #2, Johnston, IA 50131
Post Office Address *(if different)*:	same as above
Signature:	Date:
☐ A petition has been filed for this unsigned inventor.	

Name of Additional Inventor, if any:	Charles M. Papa
Citizenship:	US
Residence:	903 Beacon Street #1, Madison, WI 53715
Post Office Address *(if different)*:	1590 44th Road,Bellwood, NE 68624
Signature: Charles M. Papa	Date: 9-20-2001
☐ A petition has been filed for this unsigned inventor.	

Name of Additional Inventor, if any:	
Citizenship:	
Residence:	
Post Office Address *(if different)*:	
Signature:	Date:
☐ A petition has been filed for this unsigned inventor.	

P99172US

DECLARATION AND POWER OF ATTORNEY FOR UTILITY OR DESIGN PATENT APPLICATION (37 CFR 1.63)	Attorney Docket No.: WIS4987P0052US
	First Named Inventor: Shawn M. Kaeppler *et al.*
	COMPLETE IF KNOWN
☐ Declaration Submitted With Initial Filing ☒ Declaration Submitted After Initial Filing (surcharge (37 CFR 1.16(a)) required	Application Number: 09/914,001
	Filing Date: August 20, 2001
	Group Art Unit:
	Examiner Name:

As a below-named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name.

I believe I am the original, first, and sole inventor (if only one name is listed) or an original, first and joint inventor (if plural names are listed) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **Class II DNA Methyltransferases of Zea mays**, the specification of which:

☐ is attached hereto; or

☒ was filed on March 10, 2000 as Application Serial No. PCT/US00/06456 and was amended on _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose all information which is material to patentability as defined in 37 CFR. 1.56, including for continuation-in-part applications, material information which became available between the filing date of the prior application and the national or PCT international filing date of the continuation-in-part application.

I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or 365(b) of any foreign application(s) for patent or inventor's certificate, or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Numbers	Country	Foreign Filing Date (MM/DD/YYY)	Priority Not Claimed	Certified Copy Attached?	
				YES	NO
			☐	☐	☐
			☐	☐	☐
			☐	☐	☐

☐ Additional foreign application numbers are listed on a supplemental priority data sheet attached hereto.

I hereby claim the benefit of any United States application(s) listed below.

Application Number(s)	Filing Date	
60/123,888 60/169,858	11 March 1999 09 December 1999	☐ Additional application numbers are listed on a supplemental priority data sheet attached hereto.

The undersigned hereby authorizes the U.S. attorney(s) or agent(s) named herein to accept and follow instructions from the assignee, if any, of the undersigned or from as to any action to be taken in the Patent and Trademark Office regarding this application without direct communication between the U.S. attorney(s) or agent(s) and the undersigned. In the event of a change in the persons from whom instructions may be taken, the U.S. attorney(s) or agent(s) named herein will be so notified by the undersigned.

As a named inventor, I hereby appoint the following registered practitioner(s) to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

Lawrence J. Chapa	Reg. No. 39,135	Martin L. Katz	Reg. No. 25,011	Keith V. Rockey	Reg. No. 24,713
Randall T. Erickson	Reg. No. 33,872	Kathleen A. Lyons	Reg. No. 31,852	Thomas I. Ross	Reg. No. 29,275
Stephen D. Geimer	Reg. No. 28,846	John P. Milnamow	Reg. No. 20,635	Joel E. Siegel	Reg. No. 25,440
H. Vincent Harsha	Reg. No. 18,045	Lisa V. Mueller	Reg. No. 38,978	Paul M. Vargo	Reg. No. 29,116
Allen J. Hoover	Reg. No. 24,103	Paul M. Odell	Reg. No. 28,332		

whose mailing address for this application is: ROCKEY, MILNAMOW & KATZ, LTD.
Two Prudential Plaza - Suite 4700
180 North Stetson Avenue
Chicago, Illinois 60601
Telephone: (312) 616-5400
Facsimile: (312) 616-5460

Customer Number (01942) and/or Bar Code Label:

I hereby declare that all statements made herein of my knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Name of Sole or First Inventor:	Shawn M. Kaeppler	
Citizenship:	US	
Residence:	5290 County Highway A, Oregon, Wisconsin 53575	
Post Office Address *(if different)*:	same as above	
Signature:		Date:
☐ A petition has been filed for this unsigned inventor.		

SEQUENCE LISTING

<110> Wisconsin Alumni Research Foundation et al.
Pioneer Hi-Bred, International Inc.
Regents of The University of Minnesota
Kaeppler, Shawn M.
Springer, Nathan M.
Muszynski, Michael G.
Papa, Charles M.

<120> Nucleic Acid and Amino Acid Sequences Encoding Class II DNA Methyltransferases

<130> WIS4987P0051PCT

<140> PCT/US00/06456
<141> 2000-03-10

<160> 90

<170> PatentIn Ver. 2.1

<210> 1
<211> 2736
<212> DNA
<213> Zea mays

<400> 1
atggcgccga gctccccgtc acccgccgcg cctacacgcg tctctgggcg gaagcgcgcc 60
gccaaggccg aggagatcca ccagaacaag gaggaggagg aggaggtcgc ggcggcgtcc 120
tccgccaagc gcagccgcaa ggcggcatct tccgggaaga agcccaagtc gccccccaag 180
gaggccaagc cggggaggaa gaagaagggg gatgccgaga tgaaggagcc cgtggaggac 240
gacgtgtgcg ccgaggagcc cgacgaggag gagttggcca tgggcgagga ggaggccgag 300
gagcaggcca tgcaggagga ggtggttgcg gtcgcggcgg ggtcacccgg gaagaagagg 360
gtggggagaa ggaacgccgc cgccgccgct ggcgaccacg agccggagtt catcggcagc 420
cctgttgccg cggacgaggc gcgcagcaac tggcccaagc gctacggccg cagcactgcc 480
gcaaagaaac cggatgagga ggaagagctc aaggccagat gtcactaccg gagcgctaag 540
gtggacaacg tcgtctactg cctcggggat gacgtctatg tcaaggctgg agaaaacgag 600
gcagattaca ttggccgcat tactgaattt tttgagggga ctgaccagtg tcactatttt 660
acttgccgtt ggttcttccg agcagaggac acggttatca attctttggt gtccataagt 720
gtggatggcc acaagcatga ccctagacgt gtttttcttt ctgaggaaaa gaacgacaat 780
gtgcttgatt gcattatctc caaggtcaag atagtccatg ttgatccaaa tatggatcca 840
aaagccaagg ctcagctgat agagagttgc gacctatact atgacatgtc ttactctgtt 900
gcatattcta catttgctaa tatctcgtct gaaaatgggc agtcaggcag tgataccgct 960
tcgggtattt cttctgatga tgtggatctg gagacgtcat ctagtatgcc aacgaggaca 1020
gcaacccttc ttgatctgta ttctggctgt gggggcatgt ctactggtct ttgcttgggt 1080
gcagctcttt ctggcttgaa acttgaaact cgatgggctg ttgatttcaa cagttttgcg 1140
tgccaaagtt taaaatataa tcatccacag actgaggtgc gaaatgagaa agccgatgag 1200

tttcttgccc tccttaagga atgggcagtt ctatgcaaaa aatatgtcca agatgtggat 1260
tcaaatttag caagctcaga ggatcaagcg gatgaagaca gccctcttga caaggacgaa 1320
tttgttgtag agaagcttgt cgggatatgt tatggtggca gtgacaggga aaatggcatc 1380
tattttaagg tccagtggga aggatacggc cctgaggagg atacatggga accgattgat 1440
aacttgagtg actgcccgca gaaaattaga gaatttgtac aagaagggca caaaagaaag 1500
attctcccac tgcctggtga tgttgatgtc atttgtggag gcccaccatg ccaaggtatc 1560
agtgggttta atcggtacag aaaccgtgat gagccactca aagatgagaa aaacaaacaa 1620
atggtgactt tcatggatat tgtggcgtac ttgaagccca agtatgttct catggaaaat 1680
gtggtggaca tactcaaatt tgcggatggt tacctaggaa aatatgcttt gagctgcctt 1740
gttgctatga agtaccaagc gcggcttgga atgatggtgg ctggttgcta tggtctgcca 1800
cagttcagga tgcgtgtgtt cctctggggt gctctttctt ccatggtgct ccctaagtat 1860
cctctgccca cctatgatgt tgtagtacgt ggaggagccc ctaatgcctt ttcgcaatgt 1920
atggttgcat atgacgagac acaaaaacca tccctgaaaa aagccttgct tcttggcgat 1980
gcaatttcag atttaccaaa ggttcaaaat caccagccta acgatgtgat ggagtatggt 2040
ggttccccca agaccgaatt ccagcgctac attcgactca gtcgtaaaga catgttggat 2100
tggtccttcg gtgaggggggc tggtccagat gaaggcaagc tcttggatca ccagccttta 2160
cggcttaaca acgatgatta tgagcgggtt caacagattc ctgtcaagaa gggagccaac 2220
ttccgcgacc taaagggcgt gagggttgga gcaaacaata ttgttgagtg ggatccagaa 2280
attcgagcgtg tgaaactttc atctgggaaa ccactggttc ctgactatgc aatgtcattc 2340
attcaagggca aatcactcaa gccgtttggg cgcctgtggt gggacgagac agttcctaca 2400
gttgtaacca gagcagagcc tcacaaccag gttataattc atccgactca agcaagggtc 2460
ctcactatcc gggagaacgc aaggttacag ggcttccccg attactaccg attgtttggc 2520
ccgatcaagg agaagtacat tcaagtcggg aacgcagtgg ctgtccctgt tgcccgggca 2580
ctgggctact gtctggggca agcctacctg ggtgaatctg aggggagtga ccctctgtac 2640
cagctgcctc caagtttcac ctctgttgga ggacgcactg cggggcaggc gagggcctct 2700
cctgttggca cccctgcagg ggaggtagtt gagcag 2736

<210> 2
<211> 2796
<212> DNA
<213> Zea mays

<400> 2
agagcagcag cagctaccgc agcccctgcc atggcgccga gctccccgtc acccgccgcg 60
cctacacgcg tctctgggcg gaagcgcgcc gccaaggccg aggagatcca ccagaacaag 120
gaggaggagg aggaggtcgc ggcggcgtcc tccgccaagc gcagccgcaa ggcggcatct 180
tccgggaaga agcccaagtc gccccccaag caggccaagc cggggaggaa gaagaagggg 240
gatgccgaga tgaaggagcc cgtggaggac gacgtgtgcg ccgaggagcc cgacgaggag 300
gagttggcca tgggcgagga ggaggccgag gagcaggcca tgcaggagga ggtggttgcg 360
gtcgcggcgg ggtcacccgg gaagaagagg gtggggagaa ggaacgccgc cgccgccgct 420
ggcgaccacg agccggagtt catcggcagc cctgttgccg cggacgaggc gcgcagcaac 480
tggcccaagc gctacggccg cagcactgcc gcaaagaaac cggatgagga ggaagagctc 540
aaggccagat gtcactaccg gagcgctaag gtggacaacg tcgtctactg cctcggggat 600
gacgtctatg tcaaggctgg agaaaacgag gcagattaca ttggccgcat tactgaattt 660
tttgagggga ctgaccagtg tcactatttt acttgccgtt ggttcttccg agcagaggac 720
acggttatca attctttggt gtccataagt gtggatggcc acaagcatga ccctagacgt 780
gtttttcttt ctgaggaaaa gaacgacaat gtgcttgatt gcattatctc caaggtcaag 840

atagtccatg ttgatccaaa tatggatcca aaagccaagg ctcagctgat agagagttgc 900
gacctatact atgacatgtc ttactctgtt gcatattcta catttgctaa tatctcgtct 960
gaaaatgggc agtcaggcag tgataccgct tcgggtattt cttctgatga tgtggatctg 1020
gagacgtcat ctagtatgcc aacgaggaca gcaacccttc ttgatctgta ttctggctgt 1080
gggggcatgt ctactggtct ttgcttgggt gcagctcttt ctggcttgaa acttgaaact 1140
cgatgggctg ttgatttcaa cagttttgcg tgccaaagtt taaaatataa tcatccacag 1200
actgaggtgc gaaatgagaa agccgatgag tttcttgccc tccttaagga atgggcagtt 1260
ctatgcaaaa aatatgtcca agatgtggat tcaaatttag caagctcaga ggatcaagcg 1320
gatgaagaca gccctcttga caaggacgaa tttgttgtag agaagcttgt cgggatatgt 1380
tatggtggca gtgacaggga aaatggcatc tattttaagg tccagtggga aggatacggc 1440
cctgaggagg atacatggga accgattgat aacttgagtg actgcccgca gaaaattaga 1500
gaatttgtac aagaagggca caaaagaaag attctcccac tgcctggtga tgttgatgtc 1560
atttgtggag gcccaccatg ccaaggtatc agtgggttta atcggtacag aaaccgtgat 1620
gagccactca aagatgagaa aaacaaacaa atggtgactt tcatggatat tgtggcgtac 1680
ttgaagccca agtatgttct catggaaaat gtggtggaca tactcaaatt tgcggatggt 1740
tacctaggaa aatatgcttt gagctgcctt gttgctatga agtaccaagc gcggcttgga 1800
atgatggtgg ctggttgcta tggtctgcca vshyyvshhs yhvhyhyhyy cctctggggt 1860
ggtctttctt ccatggtgct ccctaagtat cctctgccca cctatgatgt tgtagtacgt 1920
gggaggagccc ctaatgcctt ttcgcaatgt atggttgcat atgacgagac acaaaaacca 1980
tgcctgaaaa aagccttgct tcttggcgat gcaatttcag atttaccaaa ggttcaaaat 2040
gaccagccta acgatgtgat ggagtatggt ggttccccca agaccgaatt ccagcgctac 2100
agtcgactca gtcgtaaaga catgttggat tggtccttcg gtgaggggggc tggtccagat 2160
ggaaggcaagc tcttggatca ccagccttta cggcttaaca acgatgatta tgagcgggtt 2220
ctaacagattc ctgtcaagaa gggagccaac ttccgcgacc taaagggcgt gagggttgga 2280
gcaaaacaata ttgttgagtg ggatccagaa atcgagcgtg tgaaactttc atctgggaaa 2340
acactggttc ctgactatgc aatgtcattc atcaagggca aatcactcaa gccgtttggg 2400
cgcctgtggt gggacgagac agttcctaca gttgtaacca gagcagagcc tcacaaccag 2460
gttataattc atccgactca agcaagggtc ctcactatcc gggagaacgc aaggttacag 2520
cgcttccccg attactaccg attgtttggc ccgatcaagg agaagtacat tcaagtcggg 2580
agcgcagtgg ctgtccctgt tgcccgggca ctgggctact gtctggggca agcctacctg 2640
ggtgaatctg aggggagtga ccctctgtac cagctgcctc caagtttcac ctctgttgga 2700
ggacgcactg cggggcaggc gagggcctct cctgttggca cccctgcagg ggaggtagtt 2760
gagcagtaaa aggatgacag atctgagctg agctgg 2796

<210> 3
<211> 912
<212> PRT
<213> Zea mays

<400> 3
Met Ala Pro Ser Ser Pro Ser Pro Ala Ala Pro Thr Arg Val Ser Gly
1 5 10 15

Arg Lys Arg Ala Ala Lys Ala Glu Glu Ile His Gln Asn Lys Glu Glu
20 25 30

Glu Glu Glu Val Gly Ala Ala Ala Ser Ser Ala Lys Arg Ser Arg Lys

```
          35                  40                  45

Ala Ala Ser Ser Gly Lys Lys Pro Lys Ser Pro Pro Lys Gln Ala Lys
     50                  55                  60

Pro Gly Lys Lys Lys Gly Asp Ala Glu Met Lys Glu Pro Val Glu Asp
 65                  70                  75                  80

Asp Val Cys Ala Glu Glu Pro Asp Glu Glu Glu Leu Ala Met Gly Glu
                 85                  90                  95

Glu Glu Ala Glu Glu Gln Ala Met Gln Glu Glu Val Val Ala Val Ala
            100                 105                 110

Ala Gly Ser Pro Gly Lys Lys Arg Val Gly Arg Arg Asn Ala Ala Ala
        115                 120                 125

Ala Ala Gly Asp His Glu Pro Glu Phe Ile Gly Ser Pro Val Ala Ala
    130                 135                 140

Asp Glu Ala Arg Ser Asn Trp Pro Lys Arg Tyr Gly Arg Ser Thr Ala
145                 150                 155                 160

Ala Lys Lys Pro Asp Glu Glu Glu Glu Leu Lys Ala Arg Cys His Tyr
                165                 170                 175

Arg Ser Ala Lys Val Asp Asn Val Val Tyr Cys Leu Gly Asp Asp Val
            180                 185                 190

Tyr Val Lys Ala Gly Glu Asn Glu Ala Asp Tyr Ile Gly Arg Ile Thr
        195                 200                 205

Glu Phe Phe Glu Gly Thr Asp Gln Cys His Tyr Phe Thr Cys Arg Trp
    210                 215                 220

Phe Phe Arg Ala Glu Asp Thr Val Ile Asn Ser Leu Val Ser Ile Ser
225                 230                 235                 240

Val Asp Gly His Lys His Asp Pro Arg Arg Val Phe Leu Ser Glu Glu
                245                 250                 255

Lys Asn Asp Asn Val Leu Asp Cys Ile Ile Ser Lys Val Lys Ile Val
            260                 265                 270

His Val Asp Pro Asn Met Asp Pro Lys Ala Lys Ala Gln Leu Ile Glu
        275                 280                 285

Ser Cys Asp Leu Tyr Tyr Asp Met Ser Tyr Ser Val Ala Tyr Ser Thr
```

290 295 300

Phe Ala Asn Ile Ser Ser Glu Asn Gly Gln Ser Gly Ser Asp Thr Ala
305 310 315 320

Ser Gly Ile Ser Ser Asp Asp Val Asp Leu Glu Thr Ser Ser Ser Met
325 330 335

Pro Thr Arg Thr Ala Thr Leu Leu Asp Leu Tyr Ser Gly Cys Gly Gly
340 345 350

Met Ser Thr Gly Leu Cys Leu Gly Ala Ala Leu Ser Gly Leu Lys Leu
355 360 365

Glu Thr Arg Trp Ala Val Asp Phe Asn Ser Phe Ala Cys Gln Ser Leu
370 375 380

Lys Tyr Asn His Pro Gln Thr Glu Val Arg Asn Glu Lys Ala Asp Glu
385 390 395 400

Phe Leu Ala Leu Leu Lys Glu Trp Ala Val Leu Cys Lys Lys Tyr Val
405 410 415

Gln Asp Val Asp Ser Asn Leu Ala Ser Ser Glu Asp Gln Ala Asp Glu
420 425 430

Asp Ser Pro Leu Asp Lys Asp Glu Phe Val Val Glu Lys Leu Val Gly
435 440 445

Ile Cys Tyr Gly Gly Ser Asp Arg Glu Asn Gly Ile Tyr Phe Lys Val
450 455 460

Gln Trp Glu Gly Tyr Gly Pro Glu Glu Asp Thr Trp Glu Pro Ile Asp
465 470 475 480

Asn Leu Ser Asp Cys Pro Gln Lys Ile Arg Glu Phe Val Gln Glu Gly
485 490 495

His Lys Arg Lys Ile Leu Pro Leu Pro Gly Asp Val Asp Val Ile Cys
500 505 510

Gly Gly Pro Pro Cys Gln Gly Ile Ser Gly Phe Asn Arg Tyr Arg Asn
515 520 525

Arg Asp Glu Pro Leu Lys Asp Glu Lys Asn Lys Gln Met Val Thr Phe
530 535 540

Met Asp Ile Val Ala Tyr Leu Lys Pro Lys Tyr Val Leu Met Glu Asn

545 550 555 560

Val Val Asp Ile Leu Lys Phe Ala Asp Gly Tyr Leu Gly Lys Tyr Ala
565 570 575

Leu Ser Cys Leu Val Ala Met Lys Tyr Gln Ala Arg Leu Gly Met Met
580 585 590

Val Ala Gly Cys Tyr Gly Leu Pro Gln Phe Arg Met Arg Val Phe Leu
595 600 605

Trp Gly Ala Leu Ser Ser Met Val Leu Pro Lys Tyr Pro Leu Pro Thr
610 615 620

Tyr Asp Val Val Val Arg Gly Gly Ala Pro Asn Ala Phe Ser Gln Cys
625 630 635 640

Met Val Ala Tyr Asp Glu Thr Gln Lys Pro Ser Leu Lys Lys Ala Leu
645 650 655

Leu Leu Gly Asp Ala Ile Ser Asp Leu Pro Lys Val Gln Asn His Gln
660 665 670

Pro Asn Asp Val Met Glu Tyr Gly Gly Ser Pro Lys Thr Glu Phe Gln
675 680 685

Arg Tyr Ile Arg Leu Ser Arg Lys Asp Met Leu Asp Trp Ser Phe Gly
690 695 700

Glu Gly Ala Gly Pro Asp Glu Gly Lys Leu Leu Asp His Gln Pro Leu
705 710 715 720

Arg Leu Asn Asn Asp Asp Tyr Glu Arg Val Gln Gln Ile Pro Val Lys
725 730 735

Lys Gly Ala Asn Phe Arg Asp Leu Lys Gly Val Arg Val Gly Ala Asn
740 745 750

Asn Ile Val Glu Trp Asp Pro Glu Ile Glu Arg Val Lys Leu Ser Ser
755 760 765

Gly Lys Pro Leu Val Pro Asp Tyr Ala Met Ser Phe Ile Lys Gly Lys
770 775 780

Ser Leu Lys Pro Phe Gly Arg Leu Trp Trp Asp Glu Thr Val Pro Thr
785 790 795 800

Val Val Thr Arg Ala Glu Pro His Asn Gln Val Ile Ile His Pro Thr

805 810 815

Gln Ala Arg Val Leu Thr Ile Arg Glu Asn Ala Arg Leu Gln Gly Phe
820 825 830

Pro Asp Tyr Tyr Arg Leu Phe Gly Pro Ile Lys Glu Lys Tyr Ile Gln
835 840 845

Val Gly Asn Ala Val Ala Val Pro Val Ala Arg Ala Leu Gly Tyr Cys
850 855 860

Leu Gly Gln Ala Tyr Leu Gly Glu Ser Glu Gly Ser Asp Pro Leu Tyr
865 870 875 880

Gln Leu Pro Pro Ser Phe Thr Ser Val Gly Gly Arg Thr Ala Gly Gln
885 890 895

Ala Arg Ala Ser Pro Val Gly Thr Pro Ala Gly Glu Val Val Glu Gln
900 905 910

<210> 4
<211> 922
<212> PRT
<213> Zea mays

<400> 4
Arg Ala Ala Ala Ala Thr Ala Ala Pro Ala Met Ala Pro Ser Ser Pro
1 5 10 15

Ser Pro Ala Ala Pro Thr Arg Val Ser Gly Arg Lys Arg Ala Ala Lys
20 25 30

Ala Glu Glu Ile His Gln Asn Lys Glu Glu Glu Glu Glu Val Ala Ala
35 40 45

Ala Ser Ser Ala Lys Arg Ser Arg Lys Ala Ala Ser Ser Gly Lys Lys
50 55 60

Pro Lys Ser Pro Pro Lys Gln Ala Lys Pro Gly Arg Lys Lys Lys Gly
65 70 75 80

Asp Ala Glu Met Lys Glu Pro Val Glu Asp Asp Val Cys Ala Glu Glu
85 90 95

```
Pro Asp Glu Glu Glu Leu Ala Met Gly Glu Glu Glu Ala Glu Glu Gln
            100                 105                 110

Ala Met Gln Glu Glu Val Val Ala Val Ala Ala Gly Ser Pro Gly Lys
        115                 120                 125

Lys Arg Val Gly Arg Arg Asn Ala Ala Ala Ala Ala Gly Asp His Glu
    130                 135                 140

Pro Glu Phe Ile Gly Ser Pro Val Ala Ala Asp Glu Ala Arg Ser Asn
145                 150                 155                 160

Trp Pro Lys Arg Tyr Gly Arg Ser Thr Ala Ala Lys Lys Pro Asp Glu
                165                 170                 175

Glu Glu Glu Leu Lys Ala Arg Cys His Tyr Arg Ser Ala Lys Val Asp
            180                 185                 190

Asn Val Val Tyr Cys Leu Gly Asp Asp Val Tyr Val Lys Ala Gly Glu
        195                 200                 205

Asn Glu Ala Asp Tyr Ile Gly Arg Ile Thr Glu Phe Phe Glu Gly Thr
    210                 215                 220

Asp Gln Cys His Tyr Phe Thr Cys Arg Trp Phe Phe Arg Ala Glu Asp
225                 230                 235                 240

Thr Val Ile Asn Ser Leu Val Ser Ile Ser Val Asp Gly His Lys His
                245                 250                 255

Asp Pro Arg Arg Val Phe Leu Ser Glu Glu Lys Asn Asp Asn Val Leu
            260                 265                 270

Asp Cys Ile Ile Ser Lys Val Lys Ile Val His Val Asp Pro Asn Met
        275                 280                 285

Asp Pro Lys Ala Lys Ala Gln Leu Ile Glu Ser Cys Asp Leu Tyr Tyr
    290                 295                 300

Asp Met Ser Tyr Ser Val Ala Tyr Ser Thr Phe Ala Asn Ile Ser Ser
305                 310                 315                 320

Glu Asn Gly Gln Ser Gly Ser Asp Thr Ala Ser Gly Ile Ser Ser Asp
                325                 330                 335

Asp Val Asp Leu Glu Thr Ser Ser Ser Met Pro Thr Arg Thr Ala Thr
            340                 345                 350
```

Leu Leu Asp Leu Tyr Ser Gly Cys Gly Gly Met Ser Thr Gly Leu Cys
355 360 365

Leu Gly Ala Ala Leu Ser Gly Leu Lys Leu Glu Thr Arg Trp Ala Val
370 375 380

Asp Phe Asn Ser Phe Ala Cys Gln Ser Leu Lys Tyr Asn His Pro Gln
385 390 395 400

Thr Glu Val Arg Asn Glu Lys Ala Asp Glu Phe Leu Ala Leu Leu Lys
405 410 415

Glu Trp Ala Val Leu Cys Lys Lys Tyr Val Gln Asp Val Asp Ser Asn
420 425 430

Leu Ala Ser Ser Glu Asp Gln Ala Asp Glu Asp Ser Pro Leu Asp Lys
435 440 445

Asp Glu Phe Val Val Glu Lys Leu Val Gly Ile Cys Tyr Gly Gly Ser
450 455 460

Asp Arg Glu Asn Gly Ile Tyr Phe Lys Val Gln Trp Glu Gly Tyr Gly
465 470 475 480

Pro Glu Glu Asp Thr Trp Glu Pro Ile Asp Asn Leu Ser Asp Cys Pro
485 490 495

Gln Lys Ile Arg Glu Phe Val Gln Glu Gly His Lys Arg Lys Ile Leu
500 505 510

Pro Leu Pro Gly Asp Val Asp Val Ile Cys Gly Gly Pro Pro Cys Gln
515 520 525

Gly Ile Ser Gly Phe Asn Arg Tyr Arg Asn Arg Asp Glu Pro Leu Lys
530 535 540

Asp Glu Lys Asn Lys Gln Met Val Thr Phe Met Asp Ile Val Ala Tyr
545 550 555 560

Leu Lys Pro Lys Tyr Val Leu Met Glu Asn Val Val Asp Ile Leu Lys
565 570 575

Phe Ala Asp Gly Tyr Leu Gly Lys Tyr Ala Leu Ser Cys Leu Val Ala
580 585 590

Met Lys Tyr Gln Ala Arg Leu Gly Met Met Val Ala Gly Cys Tyr Gly
595 600 605

Leu Pro Gln Phe Arg Met Arg Val Phe Leu Trp Gly Ala Leu Ser Ser
610 615 620

Met Val Leu Pro Lys Tyr Pro Leu Pro Thr Tyr Asp Val Val Val Arg
625 630 635 640

Gly Gly Ala Pro Asn Ala Phe Ser Gln Cys Met Val Ala Tyr Asp Glu
645 650 655

Thr Gln Lys Pro Ser Leu Lys Lys Ala Leu Leu Leu Gly Asp Ala Ile
660 665 670

Ser Asp Leu Pro Lys Val Gln Asn His Gln Pro Asn Asp Val Met Glu
675 680 685

Tyr Gly Gly Ser Pro Lys Thr Glu Phe Gln Arg Tyr Ile Arg Leu Ser
690 695 700

Arg Lys Asp Met Leu Asp Trp Ser Phe Gly Glu Gly Ala Gly Pro Asp
705 710 715 720

Glu Gly Lys Leu Leu Asp His Gln Pro Leu Arg Leu Asn Asn Asp Asp
725 730 735

Tyr Glu Arg Val Gln Gln Ile Pro Val Lys Lys Gly Ala Asn Phe Arg
740 745 750

Asp Leu Lys Gly Val Arg Val Gly Ala Asn Asn Ile Val Glu Trp Asp
755 760 765

Pro Glu Ile Glu Arg Val Lys Leu Ser Ser Gly Lys Pro Leu Val Pro
770 775 780

Asp Tyr Ala Met Ser Phe Ile Lys Gly Lys Ser Leu Lys Pro Phe Gly
785 790 795 800

Arg Leu Trp Trp Asp Glu Thr Val Pro Thr Val Val Thr Arg Ala Glu
805 810 815

Pro His Asn Gln Val Ile Ile His Pro Thr Gln Ala Arg Val Leu Thr
820 825 830

Ile Arg Glu Asn Ala Arg Leu Gln Gly Phe Pro Asp Tyr Tyr Arg Leu
835 840 845

Phe Gly Pro Ile Lys Glu Lys Tyr Ile Gln Val Gly Asn Ala Val Ala
850 855 860

Val Pro Val Ala Arg Ala Leu Gly Tyr Cys Leu Gly Gln Ala Tyr Leu
865 870 875 880

Gly Glu Ser Glu Gly Ser Asp Pro Leu Tyr Gln Leu Pro Pro Ser Phe
885 890 895

Thr Ser Val Gly Gly Arg Thr Ala Gly Gln Ala Arg Ala Ser Pro Val
900 905 910

Gly Thr Pro Ala Gly Glu Val Val Glu Gln
915 920

<210> 5
<211> 9
<212> PRT
<213> Zea mays

<400> 5
Lys Asp Asp Arg Ser Glu Leu Ser Trp
1 5

<210> 6
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 6
tggttgctat ggtctgccac agttcag 27

<210> 7
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 7

ccagctcagc tcagatctgt catccttt 28

<210> 8
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 8
cgaaagctaa tctacacaaa cagc 24

<210> 9
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 9
gatcctctga gcttgctaaa tttg 24

<210> 10
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 10
ctcatcttgg agtggctcat cac 23

<210> 11
<211> 22
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 11
gagcacatga gggagagtgt tg 22

<210> 12
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 12
tctctaattt tctgcgggca g 21

<210> 13
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 13
cctctgccca cctatgatgt tgta 24

<210> 14
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 14
taaagggcgt gagggttgga 20

<210> 15
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 15
tcacatttgt catggcaggt tatc 24

<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 16
ctgaggaaaa gaacgacaat gtgc 24

<210> 17
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 17
gcaatcaagc acattgtcgt tcttttcctc 30

<210> 18

<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence
was artificially synthesized based on the sequence
of Zea mays.

<400> 18
gaagaagagg gtggggagaa ggaacg 26

<210> 19
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence
was artificially synthesized based on the sequence
of Zea mays.

<400> 19
ttctttgcgg cagtgctgcg 20

<210> 20
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence
was artificially synthesized based on the sequence
of Zea mays.

<400> 20
gtattgaatt gattctcaac tagtgcac 28

<210> 21
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence

was artificially synthesized based on the sequence of Zea mays.

<400> 21
caggctcaac ggcgatg 17

<210> 22
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 22
tatgcttcat cacatagacc caagtc 26

<210> 23
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 23
gatagaccta atgccaaatg agattaag 28

<210> 24
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 24
gcgatcttca gtctccacca tc 22

<210> 25
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 25
gaagacgtgc ctccatgttt catc 24

<210> 26
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 26
gttggttctt ccgagcagag g 21

<210> 27
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 27
gactgccaca tatcttatta atcgc 25

<210> 28
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 28
gcatgtgtca gcaattgctt acattc 26

<210> 29
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 29
cctctgctcg gaagaaccaa c 21

<210> 30
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 30
ctgttcggag attcatgcat gatg 24

<210> 31
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 31

ggagaacaga atggttgatt caatgg 26

<210> 32
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 32
gcacttcact ctcctggcaa acc 23

<210> 33
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 33
cggtacgctg ctgctgctct c 21

<210> 34
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 34
ccatagcatc tcacatatcg caagg 25

<210> 35
<211> 28
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 35
ggaaagaagg cagttagttg taaatggg 28

<210> 36
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 36
ggagaagcca acgccawcgc ctcyatttcg tc 32

<210> 37
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 37
ctacaacatc atagttgggc agagg 25

<210> 38
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 38
actcactata gggctcgagc ggc 23

<210> 39
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 39
taatacgact cactataggg 20

<210> 40
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 40
gatttaggtg acactatag 19

<210> 41
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 41
gttttcccag tcacgac 17

<210> 42

<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: This sequence was artificially synthesized based on the sequence of Zea mays.

<400> 42
caggaaacag ctatgac 17

<210> 43
<211> 912
<212> PRT
<213> Zea mays

<400> 43

```
Met Ala Pro Ser Ser Pro Ser Pro Ala Ala Pro Thr Arg Val Ser Gly
 1               5                  10                  15

Arg Lys Arg Ala Ala Lys Ala Glu Glu Ile His Gln Asn Lys Glu Glu
            20                  25                  30

Glu Glu Glu Val Ala Ala Ala Ser Ser Ala Lys Arg Ser Arg Lys Ala
        35                  40                  45

Ala Ser Ser Gly Lys Lys Pro Lys Ser Pro Pro Lys Gln Ala Lys Pro
    50                  55                  60

Gly Arg Lys Lys Lys Gly Asp Ala Glu Met Lys Glu Pro Val Glu Asp
65                  70                  75                  80

Asp Val Cys Ala Glu Glu Pro Asp Glu Glu Glu Leu Ala Met Gly Glu
                85                  90                  95

Glu Glu Ala Glu Glu Gln Ala Met Gln Glu Glu Val Val Ala Val Ala
            100                 105                 110

Ala Gly Ser Pro Gly Lys Lys Arg Val Gly Arg Arg Asn Ala Ala Ala
        115                 120                 125

Ala Ala Gly Asp His Glu Pro Glu Phe Ile Gly Ser Pro Val Ala Ala
    130                 135                 140

Asp Glu Ala Arg Ser Asn Trp Pro Lys Arg Tyr Gly Arg Ser Thr Ala
145                 150                 155                 160
```

Ala Lys Lys Pro Asp Glu Glu Glu Glu Leu Lys Ala Arg Cys His Tyr
165 170 175

Arg Ser Ala Lys Val Asp Asn Val Val Tyr Cys Leu Gly Asp Asp Val
180 185 190

Tyr Tyr Lys Ala Gly Glu Asn Glu Ala Asp Tyr Ile Gly Arg Ile Thr
195 200 205

Glu Phe Phe Glu Gly Thr Asp Gln Cys His Tyr Phe Thr Cys Arg Trp
210 215 220

Phe Phe Arg Ala Glu Asp Thr Val Ile Asn Ser Leu Val Ser Ile Ser
225 230 235 240

Val Asp Gly His Lys His Asp Pro Arg Arg Val Phe Leu Ser Glu Glu
245 250 255

Lys Asn Asp Asn Val Leu Asp Cys Ile Ile Ser Lys Val Lys Ile Val
260 265 270

His Val Asp Pro Asn Met Asp Pro Lys Ala Lys Ala Gln Leu Ile Glu
275 280 285

Ser Cys Asp Leu Tyr Tyr Asp Met Ser Tyr Ser Val Ala Tyr Ser Thr
290 295 300

Phe Ala Asn Ile Ser Ser Glu Asn Gly Gln Ser Gly Ser Asp Thr Ala
305 310 315 320

Ser Gly Ile Ser Ser Asp Asp Val Asp Leu Glu Thr Ser Ser Ser Met
325 330 335

Pro Thr Arg Thr Ala Thr Leu Leu Asp Leu Tyr Ser Gly Cys Gly Gly
340 345 350

Met Ser Thr Gly Leu Cys Leu Gly Ala Ala Leu Ser Gly Leu Lys Leu
355 360 365

Glu Thr Arg Trp Ala Val Asp Phe Asn Ser Phe Ala Cys Gln Ser Leu
370 375 380

Lys Tyr Asn His Pro Gln Thr Glu Val Arg Asn Glu Lys Ala Asp Glu
385 390 395 400

Phe Leu Ala Leu Leu Lys Glu Trp Ala Val Leu Cys Lys Lys Tyr Val
405 410 415

Gln Asp Val Asp Ser Asn Leu Ala Ser Ser Glu Asp Gln Ala Asp Glu
420 425 430

Asp Ser Pro Leu Asp Lys Asp Glu Phe Val Val Glu Lys Leu Val Gly
435 440 445

Ile Cys Tyr Gly Gly Ser Asp Arg Glu Asn Gly Ile Tyr Phe Lys Val
450 455 460

Gln Trp Glu Gly Tyr Gly Pro Glu Glu Asp Thr Trp Glu Pro Ile Asp
465 470 475 480

Asn Leu Ser Asp Cys Pro Gln Lys Ile Arg Glu Phe Val Gln Glu Gly
485 490 495

His Lys Arg Lys Ile Leu Pro Leu Pro Gly Asp Val Asp Val Ile Cys
500 505 510

Gly Gly Pro Pro Cys Gln Gly Ile Ser Gly Phe Asn Arg Tyr Arg Asn
515 520 525

Arg Asp Glu Pro Leu Lys Asp Glu Lys Asn Lys Gln Met Val Thr Phe
530 535 540

Met Asp Ile Val Ala Tyr Leu Lys Pro Lys Tyr Val Leu Met Glu Asn
545 550 555 560

Val Val Asp Ile Leu Lys Phe Ala Asp Gly Tyr Leu Gly Lys Tyr Ala
565 570 575

Leu Ser Cys Leu Val Ala Met Lys Tyr Gln Ala Arg Leu Gly Met Met
580 585 590

Val Ala Gly Cys Tyr Gly Leu Pro Gln Phe Arg Met Arg Val Phe Leu
595 600 605

Trp Gly Ala Leu Ser Ser Met Val Leu Pro Lys Tyr Pro Leu Pro Thr
610 615 620

Tyr Asp Val Val Val Arg Gly Gly Ala Pro Asn Ala Phe Ser Gln Cys
625 630 635 640

Met Val Ala Tyr Asp Glu Thr Gln Lys Pro Ser Leu Lys Lys Ala Leu
645 650 655

Leu Leu Gly Asp Ala Ile Ser Asp Leu Pro Lys Val Gln Asn His Gln
660 665 670

Pro Asn Asp Val Met Glu Tyr Gly Gly Ser Pro Lys Thr Glu Phe Gln
675 680 685

Arg Tyr Ile Arg Leu Ser Arg Lys Asp Met Leu Asp Trp Ser Phe Gly
690 695 700

Glu Gly Ala Gly Pro Asp Glu Gly Lys Leu Leu Asp His Gln Pro Leu
705 710 715 720

Arg Leu Asn Asn Asp Asp Tyr Glu Arg Val Gln Gln Ile Pro Val Lys
725 730 735

Lys Gly Ala Asn Phe Arg Asp Leu Lys Gly Val Arg Val Gly Ala Asn
740 745 750

Asn Ile Val Glu Trp Asp Pro Glu Ile Glu Arg Val Lys Leu Ser Ser
755 760 765

Gly Lys Pro Leu Val Pro Asp Tyr Ala Met Ser Phe Ile Lys Gly Lys
770 775 780

Ser Leu Lys Pro Phe Gly Arg Leu Trp Trp Asp Glu Thr Val Pro Thr
785 790 795 800

Val Val Thr Arg Ala Glu Pro His Asn Gln Val Ile Ile His Pro Thr
805 810 815

Gln Ala Arg Val Leu Thr Leu Arg Glu Asn Ala Arg Leu Gln Gly Phe
820 825 830

Pro Asp Tyr Tyr Arg Leu Phe Gly Pro Ile Lys Glu Lys Tyr Ile Gln
835 840 845

Val Gly Asn Ala Val Ala Val Pro Val Ala Arg Ala Leu Gly Tyr Cys
850 855 860

Leu Gly Gln Ala Tyr Leu Gly Glu Ser Glu Gly Ser Asp Pro Leu Tyr
865 870 875 880

Gln Leu Pro Pro Ser Phe Thr Ser Val Gly Gly Arg Thr Ala Gly Gln
885 890 895

Ala Arg Ala Ser Pro Val Gly Thr Pro Ala Gly Glu Val Val Glu Gln
900 905 910

<210> 44
<211> 791
<212> PRT
<213> Arabidopsis thaliana

<400> 44
Met Ala Ala Arg Asn Lys Gln Lys Lys Arg Ala Glu Pro Glu Ser Asp
1 5 10 15

Leu Cys Phe Ala Gly Lys Pro Met Ser Val Val Glu Ser Thr Ile Arg
20 25 30

Trp Pro His Arg Tyr Gln Ser Lys Lys Thr Lys Leu Gln Ala Pro Thr
35 40 45

Lys Lys Pro Ala Asn Lys Gly Gly Lys Lys Glu Asp Glu Glu Ile Ile
50 55 60

Lys Gln Ala Lys Cys His Phe Asp Lys Ala Leu Val Asp Gly Val Leu
65 70 75 80

Ile Asn Leu Asn Asp Asp Val Tyr Val Thr Gly Leu Pro Gly Lys Leu
85 90 95

Lys Phe Ile Ala Lys Val Ile Glu Leu Phe Glu Ala Asp Asp Gly Val
100 105 110

Pro Tyr Cys Arg Phe Arg Trp Tyr Tyr Arg Pro Glu Asp Thr Leu Ile
115 120 125

Glu Arg Phe Ser His Leu Val Gln Pro Lys Arg Val Phe Leu Ser Asn
130 135 140

Asp Glu Asn Asp Asn Pro Leu Thr Cys Ile Trp Ser Lys Val Asn Ile
145 150 155 160

Ala Lys Val Pro Leu Pro Lys Ile Thr Ser Arg Ile Glu Gln Arg Val
165 170 175

Ile Pro Pro Cys Asp Tyr Tyr Tyr Asp Met Lys Tyr Glu Val Pro Tyr
180 185 190

Leu Asn Phe Thr Ser Ala Asp Asp Gly Ser Asp Ala Ser Ser Ser Leu
195 200 205

Ser Ser Asp Ser Ala Leu Asn Cys Phe Glu Asn Leu His Lys Asp Glu

210 215 220

Lys Phe Leu Leu Asp Leu Tyr Ser Gly Cys Gly Ala Met Ser Thr Gly
225 230 235 240

Phe Cys Met Gly Ala Ser Ile Ser Gly Val Lys Leu Ile Thr Lys Trp
245 250 255

Ser Val Asp Ile Asn Lys Phe Ala Cys Asp Ser Leu Lys Leu Asn His
260 265 270

Pro Glu Thr Glu Val Arg Asn Glu Ala Ala Glu Asp Phe Leu Ala Leu
275 280 285

Leu Lys Glu Trp Lys Arg Leu Cys Glu Lys Phe Ser Leu Val Ser Ser
290 295 300

Thr Glu Pro Val Glu Ser Ile Ser Glu Leu Glu Asp Glu Glu Val Glu
305 310 315 320

Glu Asn Asp Asp Ile Asp Glu Ala Ser Thr Gly Ala Glu Leu Glu Pro
325 330 335

Gly Glu Phe Glu Val Glu Lys Phe Leu Gly Ile Met Phe Gly Asp Pro
340 345 350

Gln Gly Thr Gly Glu Lys Thr Leu Gln Leu Met Val Arg Trp Lys Gly
355 360 365

Tyr Asn Ser Ser Tyr Asp Thr Trp Glu Pro Tyr Ser Gly Leu Gly Asn
370 375 380

Cys Lys Glu Lys Leu Lys Glu Tyr Val Ile Asp Gly Phe Lys Ser His
385 390 395 400

Leu Leu Pro Leu Pro Gly Thr Val Tyr Thr Val Cys Gly Gly Pro Pro
405 410 415

Cys Gln Gly Ile Ser Gly Tyr Asn Arg Tyr Arg Asn Asn Glu Ala Pro
420 425 430

Leu Glu Asp Gln Lys Asn Gln Gln Leu Leu Val Phe Leu Asp Ile Ile
435 440 445

Asp Phe Leu Lys Pro Asn Tyr Val Leu Met Glu Asn Val Val Asp Leu
450 455 460

Leu Arg Phe Ser Lys Gly Phe Leu Ala Arg His Ala Val Ala Ser Phe

465 470 475 480

Val Ala Met Asn Tyr Gln Thr Arg Leu Gly Met Met Ala Ala Gly Ser
485 490 495

Tyr Gly Leu Pro Gln Leu Arg Asn Arg Val Phe Leu Trp Ala Ala Gln
500 505 510

Pro Ser Glu Lys Leu Pro Pro Tyr Pro Leu Pro Thr His Glu Val Ala
515 520 525

Lys Lys Phe Asn Thr Pro Lys Glu Phe Lys Asp Leu Gln Val Gly Arg
530 535 540

Ile Gln Met Glu Phe Leu Lys Leu Asp Asn Ala Leu Thr Leu Ala Asp
545 550 555 560

Ala Ile Ser Asp Leu Pro Pro Val Thr Asn Tyr Val Ala Asn Asp Val
565 570 575

Met Asp Tyr Asn Asp Ala Ala Pro Lys Thr Glu Phe Glu Asn Phe Ile
580 585 590

Ser Leu Lys Arg Ser Glu Thr Leu Leu Pro Ala Cys Gly Gly Asp Pro
595 600 605

Thr Arg Arg Leu Phe Asp His Gln Pro Leu Val Leu Gly Asp Asp Asp
610 615 620

Leu Glu Arg Val Ser Tyr Ile Pro Lys Gln Lys Gly Ala Asn Tyr Arg
625 630 635 640

Asp Met Pro Gly Val Leu Val His Asn Asn Lys Ala Glu Ile Asn Pro
645 650 655

Arg Phe Arg Ala Lys Leu Lys Ser Gly Lys Asn Val Val Pro Ala Tyr
660 665 670

Ala Ile Ser Phe Ile Lys Gly Lys Ser Lys Lys Pro Phe Gly Arg Leu
675 680 685

Trp Gly Asp Glu Ile Val Asn Thr Val Val Thr Arg Ala Glu Pro His
690 695 700

Asn Gln Cys Val Ile His Pro Met Gln Asn Arg Val Leu Ser Val Arg
705 710 715 720

Glu Asn Ala Arg Leu Gln Gly Phe Pro Asp Cys Tyr Lys Leu Cys Gly

725 730 735

Thr Ile Lys Glu Lys Tyr Ile Gln Val Gly Asn Ala Val Ala Val Pro
740 745 750

Val Gly Val Ala Leu Gly Tyr Ala Phe Gly Met Ala Ser Gln Gly Leu
755 760 765

Thr Asp Asp Glu Pro Val Ile Lys Leu Pro Phe Lys Tyr Pro Glu Cys
770 775 780

Met Gln Ala Lys Asp Gln Ile
785 790

<210> 45
<211> 444
<212> PRT
<213> Zea mays

<400> 45
Leu Asp Ile Phe Ala Gly Cys Gly Gly Leu Ser Glu Gly Leu Gln Gln
1 5 10 15

Ala Gly Val Ser Phe Thr Lys Trp Ala Ile Glu Tyr Glu Glu Pro Ala
20 25 30

Gly Glu Ala Phe Asn Lys Asn His Pro Glu Ala Val Val Phe Val Asp
35 40 45

Asn Cys Asn Val Ile Leu Lys Ala Ile Met Asp Lys Cys Gly Asp Thr
50 55 60

Asp Asp Cys Val Ser Thr Ser Glu Ala Ala Glu Gln Ala Ala Lys Leu
65 70 75 80

Pro Glu Val Asn Ile Asn Asn Leu Pro Val Pro Gly Glu Val Glu Phe
85 90 95

Ile Asn Gly Gly Pro Pro Cys Gln Gly Phe Ser Gly Met Asn Arg Phe
100 105 110

Asn Cys Gln Ser Pro Trp Ser Lys Val Gln Cys Glu Met Ile Leu Ala
115 120 125

Phe Leu Ser Phe Ala Glu Tyr Phe Arg Pro Arg Phe Phe Leu Leu Glu
130 135 140

Asn Val Arg Asn Phe Val Ser Phe Asn Lys Gly Gln Thr Phe Arg Leu
145 150 155 160

Ala Val Ala Ser Leu Leu Glu Met Gly Tyr Gln Val Arg Phe Gly Ile
165 170 175

Leu Glu Ala Gly Ala Phe Gly Val Ala Gln Ser Arg Lys Arg Ala Phe
180 185 190

Ile Trp Ala Ala Ala Pro Gly Glu Met Leu Pro Asp Trp Pro Glu Pro
195 200 205

Met His Val Phe Ala Ser Pro Glu Leu Lys Ile Thr Leu Pro Asp Gly
210 215 220

Gln Tyr Tyr Ala Ala Ala Arg Ser Thr Ala Gly Gly Ala Pro Phe Arg
225 230 235 240

Ala Ile Thr Val Arg Asp Thr Ile Gly Asp Leu Pro Lys Val Gly Asn
245 250 255

Gly Ala Ser Lys Leu Thr Leu Glu Tyr Gly Gly Glu Pro Val Ser Trp
260 265 270

Phe Gln Lys Lys Ile Arg Gly Ser Met Met Val Leu Asn Asp His Ile
275 280 285

Ser Lys Glu Met Asn Glu Leu Asn Leu Ile Arg Cys Gln His Ile Pro
290 295 300

Lys Arg Pro Gly Cys Asp Trp His Asp Leu Pro Asp Glu Lys Val Lys
305 310 315 320

Leu Ser Asn Gly Gln Met Ala Asp Leu Ile Pro Trp Cys Leu Pro Asn
325 330 335

Thr Ala Lys Arg His Asn Gln Trp Lys Gly Cys Leu Tyr Gly Arg Leu
340 345 350

Asp Trp Glu Gly Asn Phe Pro Thr Ser Val Thr Asp Pro Gln Pro Met
355 360 365

Gly Lys Val Gly Met Cys Phe His Pro Asp Gln Asp Arg Ile Ile Thr
370 375 380

Val Arg Glu Cys Ala Arg Ser Gln Gly Phe Pro Asp Ser Tyr Glu Phe
385 390 395 400

Ala Gly Asn Ile Gln Asn Lys His Arg Gln Ile Gly Asn Ala Val Pro
405 410 415

Pro Pro Leu Ala Tyr Ala Leu Gly Arg Lys Leu Lys Glu Ala Val Asp
420 425 430

Lys Arg Gln Glu Ala Ser Ala Gly Val Pro Ala Pro
435 440

<210> 46
<211> 440
<212> PRT
<213> Arabidopsis thaliana

<400> 46
Leu Asp Ile Phe Ala Gly Cys Gly Gly Leu Ser His Gly Leu Lys Lys
1 5 10 15

Ala Gly Val Ser Asp Ala Lys Trp Ala Ile Glu Tyr Glu Glu Pro Ala
20 25 30

Gly Gln Ala Phe Lys Gln Asn His Pro Glu Ser Thr Val Phe Val Asp
35 40 45

Asn Cys Asn Val Ile Leu Arg Ala Ile Met Glu Lys Gly Gly Asp Gln
50 55 60

Asp Asp Cys Val Ser Thr Thr Glu Ala Asn Glu Leu Ala Ala Lys Leu
65 70 75 80

Thr Glu Glu Gln Lys Ser Thr Leu Pro Leu Pro Gly Gln Val Asp Phe
85 90 95

Ile Asn Gly Gly Pro Pro Cys Gln Gly Phe Ser Gly Met Asn Arg Phe
100 105 110

Asn Cys Gln Ser Ser Trp Ser Lys Val Gln Cys Glu Met Ile Leu Ala
115 120 125

Phe Leu Ser Phe Ala Asp Tyr Phe Arg Pro Arg Tyr Phe Leu Leu Glu
130 135 140

Asn Val Arg Thr Phe Val Ser Phe Asn Lys Gly Gln Thr Phe Gln Leu
145 150 155 160

Thr Leu Ala Ser Leu Leu Glu Met Gly Tyr Gln Val Arg Phe Gly Ile
165 170 175

Leu Glu Ala Gly Ala Tyr Gly Val Ser Gln Ser Arg Lys Arg Ala Phe
180 185 190

Ile Trp Ala Ala Ala Pro Glu Glu Val Leu Pro Glu Trp Pro Glu Pro
195 200 205

Met His Val Phe Gly Val Pro Lys Leu Lys Ile Ser Leu Ser Gln Gly
210 215 220

Leu His Tyr Ala Ala Val Arg Ser Thr Ala Leu Gly Ala Pro Phe Arg
225 230 235 240

Pro Ile Thr Val Arg Asp Thr Ile Gly Asp Leu Pro Ser Val Glu Asn
245 250 255

Gly Asp Ser Arg Thr Asn Lys Glu Tyr Lys Glu Val Ala Val Ser Trp
260 265 270

Phe Gln Lys Glu Ile Arg Gly Asn Thr Ile Ala Leu Thr Asp His Ile
275 280 285

Cys Lys Ala Met Asn Glu Leu Asn Leu Ile Arg Cys Lys Leu Ile Pro
290 295 300

Thr Arg Pro Gly Ala Asp Trp His Asp Leu Pro Lys Arg Lys Val Thr
305 310 315 320

Leu Ser Asp Gly Arg Val Glu Glu Met Ile Pro Phe Cys Leu Pro Asn
325 330 335

Thr Ala Glu Arg His Asn Gly Trp Lys Gly Leu Tyr Gly Arg Leu Asp
340 345 350

Trp Gln Gly Asn Phe Pro Thr Ser Val Thr Asp Pro Gln Pro Met Gly
355 360 365

Lys Val Gly Met Cys Phe His Pro Glu Gln His Arg Ile Leu Thr Val
370 375 380

Arg Glu Cys Ala Arg Ser Gln Gly Phe Pro Asp Ser Tyr Glu Phe Ala
385 390 395 400

Gly Asn Ile Asn His Lys His Arg Gln Ile Gly Asn Ala Val Pro Pro
405 410 415

Pro Leu Ala Phe Ala Leu Gly Arg Lys Leu Lys Glu Ala Leu His Leu
420 425 430

Lys Lys Ser Pro Gln His Gln Pro
435 440

<210> 47
<211> 130
<212> DNA
<213> Zea mays

<400> 47
catgctgttg ggccatgtgt ctagtgttgg cccattaacg tgtacacata tactagaagt 60
gtgtgtggtg tagagagagt gctgtatgtt ttccacattc cagaaaaatc cacatggtat 120
cagagccagg 130

<210> 48
<211> 123
<212> DNA
<213> Zea mays

<400> 48
gaggggagt gttgggccat gtgtctagtg ttggcccatt aacgtgtaca catatactag 60
aagtgtgtgt ggtgtagaga gagtgctgta tgttttccac attccagaaa aatccacaca 120
tgc 123

<210> 49
<211> 14
<212> PRT
<213> Zea mays

<400> 49
Cys Tyr Asn Cys Gly Asn Val Gly His Ile Ala Arg Asn Cys
1 5 10

<210> 50
<211> 17
<212> PRT
<213> Zea mays

<400> 50
Thr Gln Val Thr Gln Leu Lys Trp Ile Leu Asp Ser Gly Ala Ser Lys
1 5 10 15

His

<210> 51
<211> 14
<212> PRT
<213> Zea mays

<400> 51
Cys Gln Val Cys Ser Arg Val Gly His Thr Ala Leu Asn Cys
1 5 10

<210> 52
<211> 17
<212> PRT
<213> Zea mays

<400> 52
Gln Asn Gly Ser Asn Val Pro Trp Tyr Thr Asp Thr Gly Ala Thr Asp
1 5 10 15

His

<210> 53
<211> 14
<212> PRT
<213> Oryza sativa

<400> 53
Cys Gln Val Cys Phe Lys Arg Gly His Thr Ala Ala Asp Cys
1 5 10

<210> 54
<211> 17
<212> PRT
<213> Oryza sativa

<400> 54
Ser Tyr Gly Ile Asp Thr Asn Trp Tyr Ile Asp Thr Gly Ala Thr Asp
1 5 10 15

His

<210> 55
<211> 14
<212> PRT
<213> Arabidopsis thaliana

<400> 55
Cys Ser Asn Cys Gly Arg Thr Gly His Glu Lys Lys Glu Cys
1 5 10

<210> 56
<211> 17
<212> PRT
<213> Arabidopsis thaliana

<400> 56
Gly Lys Thr Lys Leu Gly Asp Ile Ile Leu Asp Ser Gly Ala Ser His
1 5 10 15

His

<210> 57
<211> 14
<212> PRT
<213> Zea mays

<400> 57
Cys His His Cys Gly Arg Glu Gly His Ile Lys Lys Asp Cys
1 5 10

<210> 58
<211> 17
<212> PRT
<213> Drosophila melanogaster

<400> 58
Ser Val Met Asp Asn Cys Gly Phe Val Leu Asp Ser Gly Ala Ser Asp
1 5 10 15

His

<210> 59
<211> 52

<212> PRT
<213> Zea mays

<400> 59
Gln Val Lys Ile Leu Arg Pro Asp Asn Gly Thr Glu Tyr Val Asn Lys
1 5 10 15

Gly Phe Asn Ala Phe Leu Ser Arg Asn Gly Ile Leu His Gln Thr Ser
20 25 30

Cys Pro Asp Thr Pro Pro Gln Asn Gly Val Ala Glu Arg Lys Asn Arg
35 40 45

His Ile Leu Glu
50

<210> 60
<211> 50
<212> PRT
<213> Zea mays

<400> 60
Lys Ile Ile Ala Phe Gln Ser Asp Trp Gly Gly Glu Tyr Glu Lys Leu
1 5 10 15

Asn Ala His Phe Lys Thr Ile Gly Ile His His Gln Val Ser Cys Pro
20 25 30

His Thr His Gln Gln Asn Gly Ala Ala Glu Arg Lys His Arg His Ile
35 40 45

Val Glu
50

<210> 61
<211> 51
<212> PRT
<213> Oryza sativa

<400> 61
Lys Ile Ile Ala Met Gln Thr Asp Trp Arg Gly Gly Arg Tyr Gln Lys
1 5 10 15

Leu Asn Ser Phe Phe Ala Gln Ile Gly Leu Ile Ile Met Cys His Val
20 25 30

Leu Thr Leu Ile Arg Gln Asn Gly Ser Ala Glu Arg Lys His Arg His
35 40 45

Ile Val Glu
50

<210> 62
<211> 50
<212> PRT
<213> Arabidopsis thaliana

<400> 62
Thr Val Lys Met Val Arg Ser Asp Asn Gly Thr Glu Phe Met Cys Leu
1 5 10 15

Ser Ser Tyr Phe Arg Glu Asn Gly Ile Ile His Gln Thr Ser Cys Val
20 25 30

Gly Thr Pro Gln Gln Asn Gly Arg Val Glu Arg Lys His Arg His Ile
35 40 45

Leu Asn
50

<210> 63
<211> 52
<212> PRT
<213> Drosophila melanogaster

<400> 63
Lys Val Val Tyr Leu Tyr Ile Asp Asn Gly Arg Glu Tyr Leu Ser Asn
1 5 10 15

Glu Met Arg Gln Phe Cys Val Lys Lys Gly Ile Ser Tyr His Leu Thr
20 25 30

Val Pro His Thr Pro Gln Leu Asn Gly Val Ser Glu Arg Met Ile Arg
35 40 45

Thr Ile Thr Glu
50

<210> 64
<211> 71
<212> PRT

<213> Zea mays

<400> 64
Arg Tyr Lys Ala Arg Leu Val Ala Arg Gly Tyr Ser Gln Thr Tyr Gly
1 5 10 15

Ile Asp Tyr Asp Glu Thr Phe Ala Pro Val Ala Lys Met Ser Thr Val
20 25 30

Arg Thr Leu Ile Ser Cys Ala Ala Asn Phe Gly Trp Pro Leu Tyr Gln
35 40 45

Leu Asp Val Lys Asn Ala Phe Leu His Gly Asp Leu Gln Glu Glu Val
50 55 60

Tyr Met Glu Ile Pro Pro Gly
65 70

<210> 65
<211> 12
<212> PRT
<213> Zea mays

<400> 65
Ala Ile Leu Ala Val Tyr Val Asp Asp Ile Ile Ile
1 5 10

<210> 66
<211> 71
<212> PRT
<213> Zea mays

<400> 66
Arg Leu Lys Ala Arg Leu Val Ala Lys Gly Phe Lys Gln Gln Tyr Gly
1 5 10 15

Ile Asp Tyr Asp Asp Thr Phe Ser Pro Val Val Lys His Ser Thr Ile
20 25 30

Arg Leu Val Leu Ser Leu Ala Val Ser Gln Lys Trp Ser Leu Arg Gln
35 40 45

Leu Asp Val Gln Asn Ala Phe Leu His Gly Ile Leu Glu Glu Thr Val
50 55 60

Tyr Met Lys Gln Pro Pro Gly

65 70

<210> 67
<211> 12
<212> PRT
<213> Zea mays

<400> 67
Ile Tyr Ile Leu Val Tyr Val Asp Asp Ile Ile Ile
1 5 10

<210> 68
<211> 71
<212> PRT
<213> Oryza sativa

<400> 68
Arg Tyr Lys Ala Arg Leu Val Ala Lys Gly Phe Lys Gln Arg Tyr Gly
1 5 10 15

Ile Asp Tyr Glu Asp Thr Phe Ser Pro Val Val Lys Ala Ala Thr Ile
20 25 30

Arg Ile Ile Leu Ser Ile Ala Val Ser Arg Gly Trp Ser Leu Arg Gln
35 40 45

Leu Asp Val Gln Asn Ala Phe Leu His Gly Phe Leu Glu Glu Glu Val
50 55 60

Tyr Met Gln Gln Pro Pro Gly
65 70

<210> 69
<211> 12
<212> PRT
<213> Oryza sativa

<400> 69
Met Phe Val Leu Val Tyr Val Asp Asp Ile Ile Val
1 5 10

<210> 70
<211> 71
<212> PRT

<213> Arabidopsis thaliana

<400> 70
Arg Tyr Lys Ala Arg Leu Val Val Gln Gly Asn Lys Gln Val Glu Gly
1 5 10 15

Glu Asp Tyr Lys Glu Thr Phe Ala Pro Val Val Arg Met Thr Thr Val
20 25 30

Arg Thr Leu Leu Arg Asn Val Ala Ala Asn Gln Trp Glu Val Tyr Gln
35 40 45

Met Asp Val His Asn Ala Phe Leu His Gly Asp Leu Glu Glu Glu Val
50 55 60

Tyr Met Lys Leu Pro Pro Gly
65 70

<210> 71
<211> 12
<212> PRT
<213> Arabidopsis thaliana

<400> 71
Leu Arg Val Leu Ile Tyr Val Asp Asp Leu Leu Ile
1 5 10

<210> 72
<211> 71
<212> PRT
<213> Drosophila melanogaster

<400> 72
Arg Tyr Lys Ala Arg Leu Val Ala Arg Gly Phe Thr Gln Lys Tyr Gln
1 5 10 15

Ile Asp Tyr Glu Glu Thr Phe Ala Pro Val Ala Arg Ile Ser Ser Phe
20 25 30

Arg Phe Ile Leu Ser Leu Val Ile Gln Tyr Asn Leu Lys Val His Gln
35 40 45

Met Asp Val Lys Thr Ala Phe Leu Asn Gly Thr Leu Lys Glu Glu Ile
50 55 60

Tyr Met Arg Leu Pro Gln Gly

65 70

<210> 73
<211> 12
<212> PRT
<213> Drosophila melanogaster

<400> 73
Ile Tyr Val Leu Leu Tyr Val Asp Asp Val Val Ile
1 5 10

<210> 74
<211> 62
<212> PRT
<213> Zea mays

<400> 74
Asp Ala Asp Trp Gly Ser Cys Leu Asp Asp Arg Arg Ser Thr Ser Gly
1 5 10 15

Tyr Cys Val Phe Val Gly Gly Asn Leu Val Ser Trp Arg Ser Lys Lys
20 25 30

Gln Ser Val Val Ser Arg Ser Thr Ala Glu Ala Glu Tyr Arg Ala Met
35 40 45

Ala Leu Ala Ile Cys Glu Met Leu Trp Ile Lys Gly Leu Leu
50 55 60

<210> 75
<211> 17
<212> PRT
<213> Zea mays

<400> 75
Asn Pro Val Gln His Asp Arg Thr Lys His Val Glu Ile Asp Arg Phe
1 5 10 15

Phe

<210> 76
<211> 62
<212> PRT

<213> Zea mays

<400> 76
Asp Ala Asp Trp Ala Gly Cys Pro Asp Asp Arg Lys Ser Thr Gly Gly
1 5 10 15

Tyr Ala Leu Phe Leu Gly Pro Asn Leu Ile Ser Trp Asn Ser Lys Lys
20 25 30

Gln Ser Thr Val Ser Arg Ser Ser Thr Glu Ala Glu Tyr Lys Ala Met
35 40 45

Ala Asn Ala Thr Ala Glu Val Ile Trp Leu Gln Ser Leu Leu
50 55 60

<210> 77
<211> 17
<212> PRT
<213> Zea mays

<400> 77
Lys Pro Ile Phe Asn Ala Arg Thr Lys His Ile Glu Val Asp Phe His
1 5 10 15

Phe

<210> 78
<211> 62
<212> PRT
<213> Oryza sativa

<400> 78
Asp Ala Asp Trp Ala Gly Ser Ile Asp Asp Arg Lys Ser Thr Gly Gly
1 5 10 15

Phe Ala Val Phe Leu Gly Ser Asn Leu Val Ser Trp Ser Ala Arg Lys
20 25 30

Gln Pro Thr Val Ser Arg Ser Ser Thr Glu Ala Glu Tyr Lys Ala Val
35 40 45

Ala Asn Thr Thr Ala Glu Leu Ile Trp Val Gln Thr Leu Leu
50 55 60

<210> 79
<211> 17
<212> PRT
<213> Oryza sativa

<400> 79
Asn Pro Val Phe His Ala Arg Thr Lys His Ile Glu Val Asp Tyr His
1 5 10 15

Phe

<210> 80
<211> 62
<212> PRT
<213> Arabidopsis thaliana

<400> 80
Asp Ser Asp Trp Gln Ser Cys Pro Leu Thr Arg Arg Ser Ile Ser Ala
1 5 10 15

Tyr Val Val Leu Leu Gly Gly Ser Pro Ile Ser Trp Lys Thr Lys Lys
20 25 30

Gln Asp Thr Val Ser His Ser Ser Ala Glu Ala Glu Tyr Arg Ala Met
35 40 45

Ser Tyr Ala Leu Lys Glu Ile Lys Trp Leu Arg Lys Leu Leu
50 55 60

<210> 81
<211> 17
<212> PRT
<213> Arabidopsis thaliana

<400> 81
Asn Pro Val Phe His Glu Arg Thr Lys His Ile Glu Ser Asp Cys His
1 5 10 15

Ser

<210> 82
<211> 63
<212> PRT

<213> Drosophila melanogaster

<400> 82
Asp Ser Asp Trp Ala Gly Ser Glu Ile Asp Arg Lys Ser Thr Thr Gly
1 5 10 15

Tyr Leu Phe Lys Met Phe Asp Phe Asn Leu Ile Cys Trp Asn Thr Lys
20 25 30

Arg Gln Asn Ser Val Ala Ala Ser Ser Thr Glu Ala Glu Tyr Met Ala
35 40 45

Leu Phe Glu Ala Cys Arg Glu Ala Leu Trp Leu Lys Phe Leu Leu
50 55 60

<210> 83
<211> 17
<212> PRT
<213> Drosophila melanogaster

<400> 83
Asn Pro Ser Cys His Lys Arg Ala Lys His Ile Asp Ile Lys Tyr His
1 5 10 15

Phe

<210> 84
<211> 1181
<212> DNA
<213> Zea mays

<400> 84
gggaattcga ttactcacta tagcgctcga gcggccgccc gggcaggttc gaaaaccatc 60
aacctaacga tgtaatggag tatggtggtt cccccaagac agagttccag cgctacattc 120
gacttggtcg taaagacatg ttggattggt cgtttggtga ggaggctggt ccagatgaag 180
gcaagctctt ggatcaccag cccttacggc ttaacaatga tgattatgag cgggttaagc 240
aaattcctgt caagaaggga gccaacttcc gtgacctaaa gggtgtcaag gttggagcaa 300
ataatgttgt tgagtgggat ccagaagtcg aacgtgtgta cctttcgtct gggaaaccac 360
tggttcctga ctatgcgatg tcattcatca agggcaaatc actcaagcca ttcgggcgcc 420
tgtggtggga ccagacggtt cctacagttg tgaccagagc agagcctcat aaccaggtta 480
tattgcatcc gactcaagca agagtcttga ctatccggga gaacgcaagg ttacagggct 540
tccccgatta ctaccgattg tttggaccga tcaaggagaa gtatattcaa gtcgggaacg 600
cagtggcagt ccctgttgca cgggcactgg gctactgtct gggtcaagcc tacctgggtg 660
aatctgacgg gagtcagcct ctgtaccagc tgcctgcaag ttttacctct gtggggcgaa 720
ccgcggttca ggcgaatgcc gcttctgttg gcactcctgc gggggaggta gtcgagcagt 780

aaaaggatag cggagcaacc ctggttggta ttttgattcg agcccatcca gtagcatgtt 840
taccaataaa taatcattgg tcgtgctgat tcttatggtt ggagatgaat gtatgtaggg 900
tgtactcgag ctcgagtgct tgttgtactg taggttgagg tttctcatcc attggcctgc 960
ctatttgtgg atgacgtttc atttcagatt agcaatgtgc ttatttaagg tttcgtcatg 1020
tacctgtatt ctacaatcca ctattgtttc caaagacagc atttgatcct taaaaaaaac 1080
tgtaaaaaaa aaaaaacagt gcccgaaaag ccgcaaaaaa aaaaaaaaaa aaaacctgcc 1140
cgggcggccg ctcgagccct atagtgagta atcgaattcc c 1181

<210> 85
<211> 6
<212> PRT
<213> Zea mays

<400> 85
Glu Phe Asp Tyr Ser Leu
1 5

<210> 86
<211> 252
<212> PRT
<213> Zea mays

<400> 86
Arg Ser Ser Gly Arg Pro Gly Arg Phe Glu Asn His Gln Pro Asn Asp
1 5 10 15

Val Met Glu Tyr Gly Gly Ser Pro Lys Thr Glu Phe Gln Arg Tyr Ile
20 25 30

Arg Leu Gly Arg Lys Asp Met Leu Asp Trp Ser Phe Gly Glu Glu Ala
35 40 45

Gly Pro Asp Glu Gly Lys Leu Leu Asp His Gln Pro Leu Arg Leu Asn
50 55 60

Asn Asp Asp Tyr Glu Arg Val Lys Gln Ile Pro Val Lys Lys Gly Ala
65 70 75 80

Asn Phe Arg Asp Leu Lys Gly Val Lys Val Gly Ala Asn Asn Val Val
85 90 95

Glu Trp Asp Pro Glu Val Glu Arg Val Tyr Leu Ser Ser Gly Lys Pro
100 105 110

Leu Val Pro Asp Tyr Ala Met Ser Phe Ile Lys Gly Lys Ser Leu Lys
115 120 125

```
Pro Phe Gly Arg Leu Trp Trp Asp Gln Thr Val Pro Thr Val Val Thr
    130                 135                 140

Arg Ala Glu Pro His Asn Gln Val Ile Leu His Pro Thr Gln Ala Arg
145                 150                 155                 160

Val Leu Thr Ile Arg Glu Asn Ala Arg Leu Gln Gly Phe Pro Asp Tyr
                165                 170                 175

Tyr Arg Leu Phe Gly Pro Ile Lys Glu Lys Tyr Ile Gln Val Gly Asn
            180                 185                 190

Ala Val Ala Val Pro Val Ala Arg Ala Leu Gly Tyr Cys Leu Gly Gln
        195                 200                 205

Ala Tyr Leu Gly Glu Ser Asp Gly Ser Gln Pro Leu Tyr Gln Leu Pro
    210                 215                 220

Ala Ser Phe Thr Ser Val Gly Arg Thr Ala Val Gln Ala Asn Ala Ala
225                 230                 235                 240

Ser Val Gly Thr Pro Ala Gly Glu Val Val Glu Gln
                245                 250
```

<210> 87
<211> 246
<212> PRT
<213> Zea mays

<400> 87

```
Lys Val Gln Asn His Gln Pro Asn Asp Val Met Glu Tyr Gly Gly Ser
  1               5                  10                  15

Pro Lys Thr Glu Phe Gln Arg Tyr Ile Arg Leu Ser Arg Lys Asp Met
            20                  25                  30

Leu Asp Trp Ser Phe Gly Glu Gly Ala Gly Pro Asp Glu Gly Lys Leu
        35                  40                  45

Leu Asp His Gln Pro Leu Arg Leu Asn Asn Asp Asp Tyr Glu Arg Val
    50                  55                  60

Gln Gln Ile Pro Val Lys Lys Gly Ala Asn Phe Arg Asp Leu Lys Gly
65                  70                  75                  80

Val Arg Val Gly Ala Asn Asn Ile Val Glu Trp Asp Pro Glu Ile Glu
```

85 90 95

Arg Val Lys Leu Ser Ser Gly Lys Pro Leu Val Pro Asp Tyr Ala Met
100 105 110

Ser Phe Ile Lys Gly Lys Ser Leu Lys Pro Phe Gly Arg Leu Trp Trp
115 120 125

Asp Glu Thr Val Pro Thr Val Val Thr Arg Ala Glu Pro His Asn Gln
130 135 140

Val Ile Ile His Pro Thr Gln Ala Arg Val Leu Thr Ile Arg Glu Asn
145 150 155 160

Ala Arg Leu Gln Gly Phe Pro Asp Tyr Tyr Arg Leu Phe Gly Pro Ile
165 170 175

Lys Glu Lys Tyr Ile Gln Val Gly Asn Ala Val Ala Val Pro Val Ala
180 185 190

Arg Ala Leu Gly Tyr Cys Leu Gly Gln Ala Tyr Leu Gly Glu Ser Glu
195 200 205

Gly Ser Asp Pro Leu Tyr Gln Leu Pro Pro Ser Phe Thr Ser Val Gly
210 215 220

Gly Arg Thr Ala Gly Gln Ala Arg Ala Ser Pro Val Gly Thr Pro Ala
225 230 235 240

Gly Glu Val Val Glu Gln
245

<210> 88
<211> 226
<212> PRT
<213> Zea mays

<400> 88
Asn His Gln Pro Asn Asp Val Met Glu Tyr Gly Gly Ser Pro Lys Thr
1 5 10 15

Glu Phe Gln Arg Tyr Ile Arg Leu Arg Lys Asp Met Leu Asp Trp Ser
20 25 30

Phe Gly Glu Ala Gly Pro Asp Glu Gly Lys Leu Leu Asp His Gln Pro
35 40 45

Leu Arg Leu Asn Asn Asp Asp Tyr Glu Arg Val Gln Ile Pro Val Lys
50 55 60

Lys Gly Ala Asn Phe Arg Asp Leu Lys Gly Val Val Gly Ala Asn Asn
65 70 75 80

Val Glu Trp Asp Pro Glu Glu Arg Val Leu Ser Ser Gly Lys Pro Leu
85 90 95

Val Pro Asp Tyr Ala Met Ser Phe Ile Lys Gly Lys Ser Leu Lys Pro
100 105 110

Phe Gly Arg Leu Trp Trp Asp Thr Val Pro Thr Val Val Thr Arg Ala
115 120 125

Glu Pro His Asn Gln Val Ile His Pro Thr Gln Ala Arg Val Leu Thr
130 135 140

Ile Arg Glu Asn Ala Arg Leu Gln Gly Phe Pro Asp Tyr Tyr Arg Leu
145 150 155 160

Phe Gly Pro Ile Lys Glu Lys Tyr Ile Gln Val Gly Asn Ala Val Ala
165 170 175

Val Pro Val Ala Arg Ala Leu Gly Tyr Cys Leu Gly Gln Ala Tyr Leu
180 185 190

Gly Glu Ser Gly Ser Pro Leu Tyr Gln Leu Pro Ser Phe Thr Ser Val
195 200 205

Gly Arg Thr Ala Gln Ala Ala Val Gly Thr Pro Ala Gly Glu Val Val
210 215 220

Glu Gln
225

<210> 89
<211> 245
<212> PRT
<213> Zea mays

<400> 89
Arg Phe Glu Asn His Gln Pro Asn Asp Val Met Glu Tyr Gly Gly Ser
1 5 10 15

Pro Lys Thr Glu Phe Gln Arg Tyr Ile Arg Leu Gly Arg Lys Asp Met
20 25 30

Leu Asp Trp Ser Phe Gly Glu Glu Ala Gly Pro Asp Glu Gly Lys Leu
35 40 45

Leu Asp His Gln Pro Leu Arg Leu Asn Asn Asp Asp Tyr Glu Arg Val
50 55 60

Lys Gln Ile Pro Val Lys Lys Gly Ala Asn Phe Arg Asp Leu Lys Gly
65 70 75 80

Val Lys Val Gly Ala Asn Asn Val Val Glu Trp Asp Pro Glu Val Glu
85 90 95

Arg Val Tyr Leu Ser Ser Gly Lys Pro Leu Val Pro Asp Tyr Ala Met
100 105 110

Ser Phe Ile Lys Gly Lys Ser Leu Lys Pro Phe Gly Arg Leu Trp Trp
115 120 125

Asp Gln Thr Val Pro Thr Val Val Thr Arg Ala Glu Pro His Asn Gln
130 135 140

Val Ile Leu His Pro Thr Gln Ala Arg Val Leu Thr Ile Arg Glu Asn
145 150 155 160

Ala Arg Leu Gln Gly Phe Pro Asp Tyr Tyr Arg Leu Phe Gly Pro Ile
165 170 175

Lys Glu Lys Tyr Ile Gln Val Gly Asn Ala Val Ala Val Pro Val Ala
180 185 190

Arg Ala Leu Gly Tyr Cys Leu Gly Gln Ala Tyr Leu Gly Glu Ser Asp
195 200 205

Gly Ser Gln Pro Leu Tyr Gln Leu Pro Ala Ser Phe Thr Ser Val Gly
210 215 220

Arg Thr Ala Val Gln Ala Asn Ala Ala Ser Val Gly Thr Pro Ala Gly
225 230 235 240

Glu Val Val Glu Gln
245

<210> 90
<211> 12812
<212> DNA
<213> Zea mays

<400> 90
tacatcaata aaataagggg cgccaacgca attgtccctt gtttttttcta acttaaagtt 60
caagcggcaa tgtcgcatct gatgtatgaa tatcaatttg taagtactaa acatagtctt 120
aattccaaaa ttaattacac aacaaagact aaattgtaag caaacctttc aagtctaatt 180
aattcataat tacaaatgtt attgtaacat catgttaccg aatcataaac taaccaggtt 240
cccatgtgta attagtttta taattatatt atatttaata tttgtaacta attgatgtga 300
cagtactaaa attaagcctc ttaagccaaa aaatccacat attttagatt taaaatttga 360
aaacagacgt atcggctaga agagccctgt cactgtcagc taatcaatta caagaagtgg 420
cccatactag ttccatcacc agtccagtag tccaccaccc caccctacag ctgggtcatc 480
tggcacgggt ggaggggcca acggccaaaa gcgccgcgca cttccggcgg gcaccctcgc 540
ggagtcgcgg gtgacagcga aatttcaaat ccataccctc ccgctgcaga cgggccccac 600
gccgtcaaaa tttggacgct cccgctccct cgatcttttg ggtttcgttt tcccagttcc 660
caccctctct tccaccctgc cctgtttcca gatttgaccg atcccccttcg attcgatttc 720
tacacccacg gtgtccagac tccagagcac tcactctcct ggcaaacccc tttcgtcttc 780
ccaaccctag agagcagcag cagctaccgc agcccctgcc atggcgccga gctccccgtc 840
acccgccgcg cctacacgcg tctctgggcg gaagcgcgcc gccaaggccg aggagatcca 900
ccagaacaag gaggaggagg aggaggtcgc ggcggcgtcc tccgccaagc gcagccgcaa 960
ggcggcatct tccgggaaga agcccaagtc gccccccaag caggccaagc cggggaggaa 1020
gaagaagggg gatgccgaga tgaaggagcc cgtggaggac gacgtgtgcg ccgaggagcc 1080
cgacgaggag gagttggcca tgggcgagga ggaggccgag gagcaggcca tgcaggagga 1140
ggtggttgcg gtcgcggcgg ggtcacccgg gaagaagagg gtggggagaa ggaacgccgc 1200
cgccgccgct ggcgaccacg agccggagtt catcggcagc cctgttgccg cggacgaggc 1260
acgcagcaac tggcccaaag cgctacggcc gcagcacttg ccgcaaaaga agtacattat 1320
attctcccag ctctggtttt gatttgacca gatttttact ccatgtctgt tagtacttgc 1380
gagctgagca atctgctatt tgctgattta ttgtgcgtgc agaccggatg aggaggaaga 1440
gctcaaggcc agatgtcact accggagcgc taaggtggac aacgtcgtct actgcctcgg 1500
ggatgacgtc tatgtcaagg tccttgttca tcgctttctg ttgcttctgc tctcatttat 1560
gatgtgcata tgtgtttgtt aaggaagcaa gaattgcttg atttttgttg ccgactcgca 1620
ttttccgtgac gagttctgcg tatggtcacc ggtacgtggc actgatacac aacgtggtat 1680
gctggaagtc tggtagtata ttttgcatcg accaggaggt ccagatcgat atgtgcggta 1740
tagtgcttat ttgattgcac cctgttcgga gattcatgca tgatggcgtg tttagatgac 1800
gcctcccaga cagctgcctg ccaggcagct gattctggcc caggcgtccg gaatggtgaa 1860
gttgcgctgg caagattctc aggccaccta ccaaatatgc cctggagcat attgcatgct 1920
tctttttttg ttctctttcc ttctatattt atctcattgt tagtgaagtt tcacattgca 1980
cgtgtcatgg aatatttact ttcaaatcaa cgaggagatg ctagcattga ggtgtgtgat 2040
aattattaca tactagaaga tatcgtgcat gttgccattg ggattgcgaa gaatgtggaa 2100
atgatgttgg tagcttgtat taagaggtta acagttagtg ggatgacatg actattagtt 2160
agagatgatg tggatagtaa gtgggatatg atgtagatga cttgtgtgtt gagacagaac 2220
tataacatgg agttggaaat gggagcagca tggtcaaaca taccctaaat gcctgtctct 2280
acacaatgtg gtgattggtg tatagtctgg tgttaaaagc tggatacttt gattctgttg 2340
aagattgtca cacccgaatt taaggacaaa tctagataca tctcatatgt gcaccaggat 2400
agtgtataga taccaatgtc ataatcttta ttacacgacg ataatgtctt acaaaatatc 2460
tggtgttaca agatgcacct ttcaacatgt ttattgctgc aaactgtttt aattaaacag 2520
aatgcagtgt tttgaacaaa aaaatgctgc tttatcctgc atcttgtttt gcatgtgtca 2580
gcaattgctt acattccatt atgatctctg agattcttta aatttctagc atgatgaaag 2640
tatttactaa ttcaactgaa cacaaacatt gtttgaatga acaaggcaac acggatgctt 2700
ggaataatgg ttgtgtataa tatcacttag tggttttgct ctcacaccac atctttcatg 2760

ggttctttaa taatagttac tgactttaag tttcttattc ctttttgtct atcttaggct 2820
ggagaaaacg aggcagatta cattggccgc attactgaat tttttgaggg gactgaccag 2880
tgtcactatt ttacttgccg ttggttcttc cgagcagagg acacggtgtg tatttagtat 2940
tttgtcattc tatgcatgtg tggatttttc tggaatgtgg aaaacataca gcactctctc 3000
tacaccacac acacttctag tatatgtgta cacgttaatg ggccaacact agacacatgg 3060
cccaacactc cccctcaaga tgggcgatag atatcaatca tccccatctt gctacataac 3120
acatcacact cttttactcc tatacccttа gtcaagcaat ctgctatttg acctttgag 3180
tttacatgat tcaactctaa agtaccatta tctaacttct ctttgatgaa gaatcgatca 3240
atttccacat gttttgttct atcatgttga actggattgt tagctatatt catggctgac 3300
ttattatcac accataactt cagggagtct tttcttaata cattcaactc tgataagaga 3360
ccctttatcc atagcatctc acatatcgca agggccatag ctcggtattc tgcttcggcg 3420
gtggaacggg ataccacaga ttgtttcttg cttctccatg atactaaatt tcctccaaca 3480
aacacacaat atcctgaagt tgaccttcta tcatcaaggc aactacccca gtctgcatca 3540
gagtaacctt ccacctttag atgaccatga cctttaaaga ttattccctt tccaggacaa 3600
gtcttcaagt atcgcagtat acgatacact gcatcaagat gtccacttct ggggtcatgc 3660
atatatcgac tcaccacact gactgcatat gtgatatcag gtcttgtatg gcacaagtag 3720
atgagccgtc caacaagtct ttgatacctt tctttattca caggatcacc agattcagca 3780
gataatttat gattcaagtc gataggtgtt gctacaggcc gacaccccaa catacctgtt 3840
tcatcaagta gatctaaaac atatttcctt tgggagagaa ctattccttt tggagatcga 3900
gcaatctcaa taccaagaaa gtatttgaga tgaccaagat ctttaacctc aaattcctta 3960
cttagattct tctttagaca tgcaatctca agatcggcat cacctgtaat aataatatca 4020
tccacataca cagctagaat tgcaattcgt cgtccaaagt gttgataaaa aacagtgtga 4080
tctccgttgc attgtttata tcccatgcta catattgcac gtctaaatct gtcaaaccat 4140
gctcttgggg actgcttgag accatacaat gatttttca atcgacaaac tttcccaatt 4200
gtctcaggct ttgacaatcc aggagggatc tccatataga cctcctcttg caaatcacca 4260
tgtaagaaag cattcttaac atctagttga tacaagggcc atccaaaatt tgcagcacaa 4320
gagatcaatg tccttacagt actcattttt gccactggtg caaatgtctc atcataatca 4380
attccatatg tttgactata ccctcttgca accaatcttg ctttatatcg ttctaccctt 4440
ccttctgggt tttgcttcac agtgaatacc catttacaac taactgcctt ctttccttta 4500
ggtagtttct caaattccca agtttgattt ttttctagag ctttaagctc ctccaacatt 4560
gcctcacgcc agttagaatt acattgtgct tctttccaat ctcttggaat tgctacggaa 4620
tgcaatgatg caacaaatgc tctatatgat ggtgacaaag acgcatatga gacataattg 4680
ctaatgtcat gttcatatcc ataccttgtt gggggggactc cagctttagc acgcgctcct 4740
tttcgtattg caatgggcaa atcataagtg tcataatctt cagtttctcc atgagacgtc 4800
aaaggtacat ttatagcctc taatgtgttt ggagagaact gctcagtact tgatgctgaa 4860
ttggtttcag gagcctgagg ttgcacatgg gactttcttc ttgtatatac ttcgccctta 4920
tatcgtaagt cgtctccaca agatttatta ttctcgtgac taggatgtgt ctccaattca 4980
cttggcatta cttgcatctt ttgagaagca ccaatcacca cttccatttt atttggttgt 5040
tgttccattg aatcaaccat tctgttctcc ccctctcgac tagcttcatc tgtgctagta 5100
gagacagaat caagaaaaaa atttagatct gtcttctcac catagaaagg cacagtctct 5160
ctaaatgtaa catccatgct tacaaacaaa cgtcgttcac taggactcca acacttgtat 5220
ccctttttgcc ctgcaggata tccaacaaaa atgcacttca cagcacgagg atccaacttc 5280
cccacctgag gtctatgatc tctgacaaaa catgtacatc caaaaatttt aggtggaacc 5340
acaaacttat tctcaccgag aagaatctca catggagtct tcattgcaag tatttttgaa 5400
ggagtgcgat taataagata tgtggcagtc aatacagctt cactccatag gaacttcgga 5460
acattcattg taaacatcag cgaacgagca acttccaaaa tgtgacgatt cttcctttca 5520
gccacaccat tttgtggagg tgtatcagga caggatgtct gatgtaatat accatttctt 5580
gacagaaatg cattaaatcc cttgtttaca tactcggttc cattgtctgg tcttaggatt 5640

ttgacttgag tattgaattg attctcaact agtgcacaaa aattttgaaa acacttcaat 5700
acttcatctt tatgcttcat cacatagacc caagtcattc cgagaaaaac aatcgataaa 5760
gtaacaaagt acttcatccc attaatagaa gtcacaggac atgtccaaac atcagaatga 5820
actagcacaa aaggagatat actcctgata cctcgactaa tataagatgt ccttgtgtgt 5880
tttgcaaact cacaggcatc acacaatagc ttgcttttat ccaccccact cattacatca 5940
ggaaaagctt tgcatatctt atcaaaagaa agatgcccta atctacaatg caagagcatc 6000
actgcaacct ccttctcttc cattcttgtt gccagcatag tgcatattgt accattagtc 6060
ccctcatgat ccatatacca caatccatta cgcctggtag ctgtcccaag tctcttccct 6120
gtttccctct cctgaattaa acaattatct cgatcaagaa taatacgaca atccaattga 6180
tcaaccaagg cacttagtga tatcaagttg acaggaaagg ttggcacata caaaactgat 6240
gacaacttaa tagatggagt gcattgcact gtgccaacac ccttgatggg ttgtggtgta 6300
ccatcagcag tttgtataat ttctttacgt gtgggggggat atcttatata tgatgtaaat 6360
tcactggacg tgcctgtgac atgctttgat gctcctgagt ctaaaatcca ttttaactgt 6420
gtgacctgtg tgggtacaaa agcatgagca taattacctt catcagtgta ggcgaagtgg 6480
acaaaatccc ctgtgtgaga ctcctgatct ttatctccag agatttgatt tttcttcctc 6540
aactttgttt catcttcgtg ttccataaat gtttcaagtt cttcttgtgt agttgctgca 6600
ttcgcccttg cccaactcct gcctccacga cctctgccgc ctctaggagc ccctcttcct 6660
gtcccacgat taactttgga aggcttagaa caattacgtg caatatgtcc aacattacca 6720
tcaattgtaac attctctagt atctttggtt ctcatagctg aaaacacagg atgaggcggc 6780
tgtttgagaac tttctctcat cactttgagt cttgactcct cctgggatat ggcagctatg 6840
tgcttcttgta ggctaggaag agtggattga tgaaacatgg aggcacgtct tccctcgaac 6900
ttctgagttta gccccccttag caattgaagt acacgtcttt tttccaccca tttcttcgcc 6960
tcaagcaacac actctgagtg tggtagctca ataggatcat aatgatcaac atcagcccat 7020
taaacattgta actcctgaac gtactccgcc acagatcgct ccccctgttt gatattatgg 7080
taggcgatctt cagtctccac catcaacata acatttccag ctcccgagta catttcttca 7140
tagtgctttcc acatttctgc agcacttatg attgtatcaa cagtgctagc aattgctgga 7200
tatcatagaac tcaacatcca cgctgccact aaagagttta tagcatccca gtctttccat 7260
ttcatcactta acttatcctt gggctcaacg acatctcctt taacatagcc ctcgagtctc 7320
tttgccttca ataatcgcaa tgctcttcta gaccatgcca aataattttt caccccttct 7380
taacttaatct catttggcat taggtctatc ttctgaactg gttctatatg agcaacattg 7440
ttctttaattg atgatggagc ctcatccctt tttgctgaca gtaattcgac caatttacca 7500
tagaacctgct caattccttg attttccccc ataatatata tgtaataata actaggaaac 7560
tacttgggca gctgcgtcaa gatctgggtc acaacgtcta gaagccagga ccaggagcgc 7620
ctcctcttcc tcctcctccg agctggatgg atctcagtca caggacgcgg gcagcagggg 7680
ggagcagcag cacctgtgtg ccggcagctt cctcaagggt tggacgagct gcggcagctg 7740
gagagcctcc caagcacccc tatctccaga tccttgtcgc cgacgagtgc ccgcgtccac 7800
gtccttggcc gccctcgcct tgtcggcggt ggcgtcctct gtgctgtggc tcgggacctg 7860
tccctggcct cctgcgcggc ctccctgctg gcgtcggtgt actcgcccgt cttctgcctg 7920
gtcacgtccc tcgcctccct cgatcgctcg tgtgcctcgg cggcctcctt cggccgtcgc 7980
tgatctcctt ctcggtggtc ttctccgtcg aggccgaaga cactcgtcac cgcgacgcca 8040
tcgccgttga gcctggctct gataccatgt ggatttttct ggaatgtgga aaacatacag 8100
cactctctct acaccacaca cacttctagt atatgtgtac acgttaatgg gccaacacta 8160
gacacatggc ccaacagcat gtcaagtggc atagcactca catttgtcat ggcaggttat 8220
caattctttg gtgtccataa gtgtggatgg ccacaagcat gaccctagac gtgtttttct 8280
ttctgaggaa aagaacgaca atgtgcttga ttgcattatc tccaaggtca agatagtcca 8340
tgttgatcca aatgtaagtt tgctgcagtt tgctgagagc tttgtggttt tgctatacac 8400
ataatgtttc tgactaccat tgttttgttg cctacttgcc ttagatggat ccaaaagcca 8460
aggctcagct gatagagagt tgcgacctat actatgacat gtcttactct gttgcatatt 8520

ctacatttgc taatatctcg tctggtaatt ccttctgcat catctttttt ggttgactag 8580
ctgaatgcag ttagctttgc caaagagtta aatacatgag ttgttcctgc actcgaaaag 8640
ggatgtcaat aatgtccaca aactctgaaa atgtattttt agatacttaa cttgttaagt 8700
cagtaaaacc tgtcagatac ttgggttttg ggtacgatta ccatccttat gtgagtaaaa 8760
ctcgtcaagg gatgtcaatg acgtgttgat tgtgtattag atattctgtt tgttcgaaag 8820
ctaatctaca caaacagctt atgtaatgta aaacctcaaa caaacttgcc tcttcataag 8880
cttaggttta taggattagc gtttagtgca tgtaaggcct atttgcttca cggcctccct 8940
gccgagctcc tggctagaca gccatcctgg ccgtaggtgc ccgaaatcga acacctggga 9000
gccacgtttg cactagcagg ttttcctggg tgcaaaccaa acacgcctat agtgttcaag 9060
tataactgaa ttggtgctca cctttgtcta atagcttaag tttttggttt tcatcggtgc 9120
atgcaactcc atactcaata gtcaatatga tatagtgttc aagcatagaa ctctcgagtt 9180
tgaatcctgg caggggcaat caaataaaat aattgcagct tacccctatt tctacgtttg 9240
agcacatgag ggagagtgtt gaattataag tgtgttctcc atctttctct aacagatgaa 9300
ctggtttgtg catgtaactc aatatgatat ttgagtcaaa tgtttacttt aaaatcatag 9360
ttgatgcaat ttaataacat attttttgg tctcgtgtga gggagtgtac gtataactga 9420
attgcacaca tttccttata gcttaggttt ttgactgcaa ctgttggtgc atgtagctca 9480
ataactaaag ttgatctgga cagtctacag tgaataagtt tgacacttgt aaaatgtgca 9540
tgtatttta caaacgctgg cactttttc ctaatagaaa atgggcagtc aggcagtgat 9600
accgcttcgg gtatttcttc tgatgatgtg gatctggaga cgtcatctag tatgccaacg 9660
aggacagcaa cccttcttga tctgtattct ggctgtgggg gcatgtctac tggtctttgc 9720
ttgggtgcag ctctttctgg cttgaaactt gaaactgtaa tcttctaact agtcatctgt 9780
tggatagaa atgttcacga tctcagaact tattctattg ttctggcttg cagcgatggg 9840
ctgttgattt caacagtttt gcgtgccaaa gtttaaaata taatcatcca cagactgagg 9900
tatggatagt aaacttcatc ttggattcca tctgttctgt cagctactct tacaaagtgt 9960
ctggattttt ggatgtaggt gcgaaatgag aaagccgatg agtttcttgc cctccttaag 10020
gaatgggcag ttctatgcaa aaaatatgtc caagatgtgg attcaaattt agcaagctca 10080
gaggatcaag cggatgaaga cagccctctt gacaaggacg aatttgttgt agagaagctt 10140
gtcgggatat gttatggtgg cagtgacagg gaaaatggca tctattttaa ggtacttcag 10200
tgtcatttgt tcatttctac ttgattccaa caaaaaaatc aattacttaa gcctgtcaaa 10260
cgatggatat ttctgtatat tttgctgtaa cgctagattt ctgcaggtcc agtgggaagg 10320
atacggccct gaggaggata catgggaacc gattgataac ttgaggttag tgtatggtat 10380
attcgtctgct tgttgccttg tatacctatt tgcatctaat ccttgttttg caaaccagtg 10440
actgcccgca gaaaattaga gaatttgtac aagaagggca caaaagaaag attctcccac 10500
tgcctgtgag tatttagttc gttgtgattt tgctcgctat ttgtttagct ccccttttt 10560
atttggtgat atctgcctat tttattcttt caaagggtga tgttgatgtc atttgtggag 10620
gcccaccatg ccaaggtatc agtgggttta atcggtacag aaaccgtgat gagccactca 10680
aagatgagaa aaacaaacaa atggtgactt tcatggatat tgtggcgtac ttgaagccca 10740
agtatgttct catggaaaat gtggtggaca tactcaaatt tgcggatggt tacctaggaa 10800
aatatgcttt gagctgcctt gttgctatga agtaccaagc gcggcttgga atgatggtgg 10860
ctggttgcta tggtctgcca cagttcagga tgcgtgtgta cctctggggt gctctttctt 10920
ccatggtctg ttctgtacct tgctgtttta tatgcttcgc tagattcata ttgcactgtt 10980
ggctgctggc taaccaggtg tacgtgtatt tgacaattta ggtgctccct aagtatcctc 11040
tgcccaacta tgatgttgta gtacgtggag gagcccctaa tgcctttcgg tgagtgcaat 11100
cacaaaccac tactatgaaa tcatgtggaa tgtgtaaaat acgctgacca actgaatttg 11160
ttgcagcaat gtatggttgc atatgacgag acacaaaaac catccctgaa aaaagccttg 11220
cttcttggcg atgcaatttc agatttacca aaggcaagtg ttctgtcaag ttcatgcatt 11280
tctcagtgag catgctattt aactcttctc tgcaggttca aaatcaccag cctaacgatg 11340
tgatggagta tggtggttcc ccaagaccga attccagcgc tacattcgac tcagtcgtaa 11400

aggtaaaaaa ccccgtgaac tactactggt tggccttcac tacgaatatg ttaggattta 11460
atttcagaag aaccgccttt tttttcttgg tgcttcggta ctactgcagc aagctcactc 11520
ttattatcat gtcagacatg ttggattggt ccttcggtga agggggctgg tccagatgaa 11580
ggcaagctct tggatcacca gcctttacgg cttaacaacg atgattatga gcgggttcaa 11640
cagattcctg tcaagaaggt tggtggcttg gtcgcatttg tgccttcctt tgttgttttt 11700
tcccccttctg aaacaatcat ctctctttcc tatgacaggg agccaacttc cgcgacctaa 11760
agggcgtgag ggttggagca aacaatattg ttgagtggga tccagaaatc gagcgtgtga 11820
aactttcatc tgggaaacca ctggtatgtg tgctatttcc gtgctgttgt ttcctataac 11880
tgtgcaacat ttactttccc atattcaaac tcataactga cgagatgctg caactactgt 11940
aagattcatg gctaacccat gacaacattt tgcacacatc tttgttatct aggttcctga 12000
ctatgcaatg tcattcatca agggcaaatc actcaagtaa gtttcaaaac atttttttg 12060
ttttttgggg gaaaagtagg ttattgttta cttgtgctta catatgatgt tgcaggccgt 12120
ttgggcgcct gtggtgggac aagacagttc ctacagttgt aaccagagca gagcctcaca 12180
accaggtcag cttcagaaag gccactcctt ttcgccaatc cctgcatctg tatttactat 12240
tagcgtgtgt tcccatatga tcattaccga acatgttgtc cacacaggtt ataattcatc 12300
cgactcaagc aagggtcctc actatccggg agaacgcaag gttacagggc ttccccgatt 12360
attaccgatt gtttggcccg atcaaggaga agtaagttcc tgttttcaag ttgcctgtac 12420
cagatctagt cactattgaa agttttcagc agcaagccat tcatcagtta gttacagctc 12480
ttgaaagcct tacctctgaa catgtgtgct ttctctgatg gtgataggta cattcaagtc 12540
gggaacgcag tggctgtccc tgttgcccgg gcactgggct actgtctggg gcaagcctac 12600
ctgggtgaat ctgaggggag tgaccctctg taccagctgc ctccaagttt cacctctgtt 12660
ggaggacgca ctgcggggca ggcgagggcc tcttcctgtt ggcacccctg caggggaggt 12720
agttgagcag taaaaggatg acagatctga gctgagctgg gcaacatcca gcggcaggag 12780
caatttctggt tcggttcgat tcgggctcac ga 12812